Exploitation of bacterial artificial chromosome (BAC) libraries to enhance the efficiency of genome mapping.

by

Mikiko Miyagi
B.Sc (Life Sciences)

A thesis submitted for the degree of Masters by Research
to
The Queensland University of Technology,
School of Life Sciences

August 2002

Key words: anchored BAC – bacterial artificial chromosome (BAC) library – bruchid – genome mapping – large inserts – linkage map – microsatellites – molecular markers – mungbean – polymorphism – powdery mildew – restriction fragment length polymorphism (RFLP) – SSR – STS – target genome region – *Vigna radiata*

ABSTRACT

The power of molecular markers and linkage maps in genetic studies and breeding programs has long been recognised. International efforts over the last two decades or so have produced linkage maps for many species of agronomic importance, and molecular markers linked to a wide array of traits in many species have been developed. However, linkage maps produced with the current techniques often contain gaps. These gaps make gene mapping inefficient because they do not allow a whole genome scan to detect gene locations. Similarly, molecular markers have still not been widely utilised in breeding programs. This is partially due to the fact that a large number of genotypes usually need to be screened in a breeding program. This study was aimed at addressing some of these difficulties in genome mapping and marker isolation by exploiting bacterial artificial chromosome (BAC) technologies.

The first objective of this study was to construct a mungbean BAC library. DNA from two different mungbean genotypes, ATF 3640 and ACC41, was used for a BAC library construction. In total 18,816 BAC clones were picked into 49 384-well microtiter plates. The average insert size of these clones is 107 kbp. Because the genome size of mungbean is 579 Mbp, the 18,816 BAC clones represent about 3.56 genome equivalents. This translates into a 95 percent probability of recovering any specific sequence of interest from these BAC clones.

The second objective of the study was to demonstrate the feasibility of isolating polymorphic markers for gaps in linkage maps. For this study, a gap near a major locus conditioning powdery mildew (pm) resistance was selected. To develop polymorphic markers for this gap, two RFLP markers from an existing map were selected. These two markers could be mapped to this region in a reference, but none of them were able to detect differences between the pm parents. Polymorphic markers for the pm populations were developed from the BAC clones isolated from both of the RFLP markers and were successfully

mapped into the targeted gap. This new approach thus shows great promise for filling gaps quickly, which will lead to more efficient mapping of genes of importance in any species.

The third objective of the study was to demonstrate the feasibility of developing locus-specific and PCR-based SSR and STS markers for targeted traits. For this work, an RFLP marker, mgM213, closely linked to a major locus conditioning bruchid resistance, was selected. Four positive BAC clones were identified by this RFLP marker. Screening of the subclones from the four BAC clones identified one unique SSR sequence. Three sets of STS primers were also designed from sequences of the subclones. These four PCR-based markers would be able to facilitate the incorporation of bruchid resistance into breeding programs of mungbean and related species.

Finally, a set of BAC clones covering the mungbean genome was isolated. As demonstrated in this study, many markers can be generated from a single BAC clone and polymorphic markers for a new population can be easily generated from BAC subclones. Thus, with the availability of these BAC clones, it would be possible to efficiently generate framework maps for new populations of mungbean and related species.

The results from this study showed clearly that the application of BAC technology can dramatically enhance the efficiency of genome mapping. BAC technology also allows for more efficient exploitation of existing linkage maps. By combining the BAC libraries and existing linkage maps that are available for many species, we could speed up the efforts of gene tagging, and further improve our ability to breed better varieties more quickly.

Table of contents

List of Tables

List of Figures

The work contained in this thesis has not been previously submitted for a degree or diploma at any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Mikiko Miyagi

Acknowledgements

I would like to express my appreciation to my supervisors Drs. Chunji Liu and Marion Bateson for their invaluable assistance, enthusiasm and encouragement throughout the course of this research. I sincerely thank Dr. Chunji Liu for his effort in planning and providing me with an invaluable opportunity to work on this project. I would like to express my gratitude for his constant guidance, support and critical appraisal of my research. I sincerely thank Dr. Marion Bateson for her constant encouragement, guidance and support.

I am grateful to my colleague Matthew Humphry for his assistance with laboratory procedures, computer techniques and unpublished results. I thank Drs. Yuezhi Tao and Ray Shorter for their initial help and support. Also thanks to Drs. Lynne McIntyre, Chris Lambrides, and John Manners and to Janneke Drenth, Adele Schmidt and Kim Ritter for their support and encouragement. My appreciation is also due to Tom Magner for being a good friend and constant encouragement and for his technical support. Special appreciation to all the wonderful people I met at CSIRO.

Special gratitude to my partner Jeff Makinson for his constant love, encouragement and support.

Finally, I would like to express my gratitude to my parents, Dr. Hiroshi Miyagi and Mrs. Fumiko Miyagi for their invaluable support, love, understanding and encouragement in pursuing my studies.

Chapter 1

Introduction and Literature Review

1.1 Introduction

The discovery of molecular markers about twenty years ago made it possible to construct linkage maps for any organism. To date, many different marker types have been invented, and one or more molecular marker-based linkage maps have been constructed for all major species. These linkage maps have been extensively used in many research fields including gene tagging (Eagles *et al.* 2001), map based gene cloning (Bent *et al.* 1994; Martin *et al.* 1993; Yoshimura *et al.* 1996), comparative mapping (Bonierbale *et al.* 1988; Gebhardt *et al.* 1991; Menancio-Hautea *et al.* 1993a; Tanksley *et al.* 1988), and evolutionary studies (Moore *et al.* 1995). These studies have dramatically improved our knowledge in these fields and have also offered new possibilities to speed up breeding programs. However, linkage maps generated to date are mainly from random markers, and marker clusters and gaps are common features in the current linkage maps (Chalmers *et al.* 2001). These marker-void regions make many marker projects difficult. They do not allow the detection of loci in these genome regions, let alone to isolate markers closely linked to them. Thus, methods that allow the efficient generation of polymorphism markers for targeted chromosome regions could not only dramatically improve the efficiency of a gene mapping project, but also enhance the feasibility of marker-assisted selection by developing user-friendly markers.

Currently, bulked segregant analysis (BSA) is widely used for filling gaps (Lefebvre and Chèvre 1995), and there are numerous examples of success,

including identifying male-sterility genes in rice (Zhang *et al.* 1994) and *Rhynchosporium* resistance genes in barley (Barua *et al.* 1993). However, this method requires markers flanking a target, thus is not suitable for those gaps located at the ends of linkage maps or where there are no markers available for large sections of chromosomes (e.g. Campbell *et al.* 2001). Further, the BSA is based on random markers so it could be time-consuming to isolate markers for small gaps.

The last few years have seen dramatic developments in the construction and exploitation of BAC libraries. As a result, one or more BAC libraries have been constructed for many species of agronomic importance including *Arabidopsis* (Choi *et al.* 1995; Mozo *et al.* 1998), rice (Wang *et al.* 1995; Zhang *et al.* 1996a), soybean (Marek and Shoemaker 1997), sorghum (Woo *et al.* 1994) and sugarcane (Tomkins *et al.* 1999b). These BAC libraries have been extensively exploited in physical mapping, genome sequencing, as well as in generating molecular markers (Cregan *et al.* 1994). The aim of this study was to test the feasibility of exploiting BAC technologies in enhancing genome mapping and marker development using mungbean as a test species.

1.2 Genetic Markers

The idea of genetic markers is not new. In fact, the very experiments in which Mendel established the basic laws of genetics were based on phenotypic markers. However, it was not until the invention of RFLP (restriction fragment length polymorphisms) in the early 1980s that it became possible to generate markers for a whole genome. Over the last twenty years, many marker systems have been developed (Table 1.1), each with its own advantages and disadvantages.

Table 1.1: Different marker systems available

morphological markers

(e.g. seed colour, plant height, awn length)

cytological markers

(i) pachytene analysis

(ii) chromosome banding

- giemsa staining
- acid, alkali or high temperature treatment
- quinacrine mustard staining

(iii) *in situ* hybridisation (ISH)

- fluorescent *in situ* hybridisation (FISH)
- primed *in situ* hybridisation (PRINS)
- genomic *in situ* hybridisation (GISH)

molecular markers

(i) biochemical markers

- isozymes
- seed storage proteins

(ii) DNA markers

- Southern hybridisation-based
  - restriction fragment length polymorphism (RFLP)
  - variable number of tandem repeats (VNTR)
- PCR-based
  - sequence tagged sites (STS)
  - simple sequence repeats (SSR)
  - random amplified polymorphic DNA (RAPD)
  - amplified fragment length polymorphism (AFLP)
  - sequence characterised amplified region (SCAR)
  - single strand confirmation polymorphism (SSCP)
  - arbitrary primed PCR (AP-PCR)
  - DNA amplification fingerprinting (DAF)
  - inter-simple sequence repeat amplification (ISSR)
  - sequence polymorphic locus amplification test (SPLAT)
  - single nucleotide polymorphisms (SNP)

1.2.1 Morphological Markers

Morphological traits including plant height, seed shape and colour, and length of awns can be monitored visually without specialised biochemical or molecular techniques. These characters can be used as genetic markers, provided their expression is reproducible over a range of environments (Kumer 1999). Morphological markers have been effectively utilised in a number of species including maize (Edwards *et al.* 1987; Stuber *et al.* 1987), barley (Qualset *et al.* 1965) and sorghum (Tao *et al.* 1998). While morphological markers can be scored easily without the use of complex and expensive equipment, there are serious limitations associated with their use (Stuber 1992; Tanksley *et al.* 1989; Paterson *et al.* 1991). These limitations include: (i) many morphological markers have large effects on phenotype thus limiting their use in plant breeding programs, (ii) the allelic frequencies of loci mapped by morphological markers are very few, (iii) in most cases their alleles interact in a dominant-recessive manner, making it impossible to distinguish heterozygous individuals from homozygous individuals, (iv) many morphological markers express their characteristics at the whole plant level, and (v) morphological markers can be altered by epistatic and pleiotropic interactions.

1.2.2 Cytological Markers

Cytological markers are reliable and they provide excellent information on the overall organisation of plant genomes. These markers serve as an important link between classical cytogenetics and molecular approaches to genome analysis (Abbo *et al.* 1993). There are three commonly used cytological markers: pachytene analysis, chromosome banding, and *in situ* hybridisation.

1.2.2.1 Pachytene Analysis

McClintock (1929) pioneered the study of meiotic pachytene chromosomes, which provides an excellent technique (pachytene analysis) for the study of chromosome morphology. Pachytene analysis has facilitated the identification and location of centromeres, telomeres, satellites, nucleolar organising regions, heterochromatine and euchromatine in maize (*Zea mays* L). These studies enabled construction of an ideogram for each chromosome.

1.2.2.2 Chromosome Banding

Chromosome banding involves staining of mitotic metaphase chromosomes with sophisticated staining procedures such as giemsa staining; acid, alkali or high temperature treatment; or quinacrine mustard staining. These differential staining procedures detect banding patterns and facilitate the identification of the heterochromatin (darkly stained) from euchromatin regions (lightly stained) (Schulz-Schaeffer 1980).

1.2.2.3 *In situ* Hybridisation (ISH)

In situ hybridisation (ISH) involves the hybridisation of a radiolabelled probe to the same chromosome spread preparation used for karyotypes and banding pattern assays (Sessions 1990). It is used to detect the presence and chromosomal distribution of target sequences. While banding assays are only appropriate with chromosomes containing specific sequences, a wide range of sequences can be detected by ISH. ISH has many applications for genome analysis. These include the evolutionary study of species (Jellen *et al.* 1994) or of chromosomes, by painting the genome of one species onto chromosomes of another, or by

painting a genome with one of its own chromosomes (Vega *et al.* 1994), respectively. ISH also has applications in the characterisation of the chromosomal distribution of interspersed repeats (Aledo *et al.* 1995). Species specific repetitive sequences from rye have been used to identify rye chromosomes and chromosome segments in hybrids with wheat using silver stained probes (Laptian *et al.* 1986). Also, tandem arrays of rRNA gene subunits have been localised on wheat chromosome spreads using this approach (Mukai *et al.* 1991).

Two modifications of the basic ISH have been developed. They are fluorescent *in situ* hybridisation (FISH), where fluorescent dyes are used to label probes instead of radio-isotopes (Trask 1991), and primed *in situ* hybridisation (PRINS), in which target sequences are amplified directly by the polymerase chain reaction (PCR) from the DNA template (Koch *et al.* 1989). Several probes are used in FISH to hybridise to the same chromosome preparation. Each probe in FISH is labelled to give a different colour when illuminated by specific wavelengths.

Genomic *in situ* hybridisation (GISH) is another powerful technique, which has been used for the analysis of genome structure in allopolyploid species, including crops such as oilseed rape (canola), oat, wheat, groundnut, upland and sea island cotton, bananas, strawberry, arabica coffee and tobacco (Laurie *et al.* 1997). In GISH, genomic DNA from one of the progenitors of an allopolyploid species is labelled as a probe, and genomic DNA of the second progenitor, which is either differently labelled or unlabelled, is added into the hybridisation mixture during *in situ* hybridisation (King *et al.* 1993). Wheat chromosomes could be clearly differentiated from those of rye (Le *et al.* 1989), and barley (Mukai and Gill 1991) in hybrids and the three genomes could be differentiated in hexaploid wheat (Mukai *et al.* 1993).

1.2.3 Molecular Markers

Molecular markers reveal polymorphism at the protein or DNA level. Biochemical markers are able to reveal polymorphism at the protein level, while DNA markers reveal polymorphism at the DNA level.

1.2.3.1 Biochemical Markers

Biochemical markers include a wide range of gene products. The most commonly used protein markers are isozymes (e.g. esterases and peroxidases) and seed storage proteins (e.g. gliadins and glutenins). Markert and Moller (1959) exploited the efficacy of these markers when they found genetic differences that created variants of particular enzymes that could cause differences in their mobility under an electric field. The general procedure for analysing biochemical markers involves extracting total proteins from the tissues, separating proteins by gel electrophoresis, and staining in enzyme specific stains. However, the procedure varies depending on the nature of proteins to be analysed and the types of electrophoresis used.

Even though there is a long history of separating proteins by electrophoresis, substantial variation between varieties in a wide range of proteins was not achieved until acrylamide was used for the gel matrix. Polyacrylamide gel electrophoresis (PAGE) is highly adaptable and can be made to stringent specifications. Additionally, the gels possess other important qualities, such as being chemically inert, while mechanically strong, and being able to offer an exceptionally high resolution.

Isoelectric focusing (IEF) is another technique offering even greater resolution. This technique involves proteins being placed in a pH gradient across which a voltage is applied, so that they migrate to the pH at which

there is no net charge, called the isoelectric point (pI), and then cease to move. With the application of IEF in plant isozyme analyses, the number of biochemical loci identified, especially polymorphic ones, has rapidly increased (Liu 1997). Maps based on biochemical markers have been published for tomato (Tanksley and Rick 1980), maize (Goodman and Stuber 1983), and wheat (Hart 1983).

Compared with morphological markers, biochemical markers offer significant advantages in genome analysis. They include:

(1) lack of environmental effects, except when the expression of genes is measured by enzymes;

(2) high heritability and a lack of epistatic effects;

(3) quick analysis time;

(4) the ability to be detected at any growth stage of the plant, reducing the number of plants required to be grown to maturity (except when enzymes are assayed to study gene expression);

(5) the need for only a small amount of plant tissue;

(6) functioning as codominant markers and ability to be analysed as codominant variants within plant populations, species, and genera for gene mapping and variety identification (Stuber 1992; Tanksley 1983; Westman and Kresovich 1997).

While analysis of their segregation is relatively easy, several drawbacks should be noted in regard to the use of isozymes. First, the number of isozymes that can be scored is limited. Relatively few loci occur in any species for these markers (Vodenicharova 1989). A second drawback is in tissue variability. While some proteins can only be detected in a certain plant tissue, some isozymes are better expressed in certain tissue, e.g. root versus leaf tissue (Lefebvre and Chèvre 1995). Therefore, several samples of the segregating population may be necessary to score all of the available isozymes.

Immunological properties of proteins have also been used as a marker assay. Monoclonal antibodies to seed proteins of several cereal species have been produced to identify cDNA library clones and to describe variation between crop species and cultivars. Enzyme-linked immunosorbent assays (ELISA) are used to evaluate reactivity of these antibodies with antigenic proteins (Westman and Kresovich 1997).

1.2.3.2 DNA Markers

The major advantage of DNA markers over biochemical and morphological markers is their availability in unlimited number. In any genome, the number of morphological and biochemical markers is limited compared to DNA markers, which are ubiquitous and numerous. Also, they are different from morphological and biochemical markers that depend upon the expression of certain genes, which might be governed by environmental conditions or tissue specificity. DNA markers are neutral, have no effect on the phenotype, and are free of pleiotropic effects. DNA marker analysis can be carried out at any stage of the life cycle of an organism. They can even be analysed using herbarium and mummified tissue (Kumer 1999). The availability of the DNA marker system has enabled researchers to construct genetic maps, which are usually constructed to locate genes of biological or economic interest in relation to molecular markers. DNA markers reveal polymorphisms at the DNA level and can be classified into two categories: Southern hybridisation-based markers and PCR-based markers.

A. Southern hybridisation-based markers

Southern hybridisation-based markers include restriction fragment length polymorphisms (RFLP) and variable number of tandem repeats (VNTR). The polymorphisms in the case of RFLP are generated due to

events such as point mutations at restriction enzyme recognition sites, insertions, deletions, or translocations of sequences at, or between, restriction enzyme recognition sites (Figure 1.1). VNTR loci are due to a difference in the number of repeats (Kumer 1999).

- RFLP

The use of RFLP as genetic markers was first proposed by Botstein *et al.* (1980). Restriction enzymes were used to identify single base-pair changes in genomic DNA that result in the gain or loss of a restriction site. Variations in these nucleotides were called 'restriction fragment length polymorphisms (RFLP)' and were used in early linkage studies.

Probes used for RFLP analysis can be derived from random pieces of genomic DNA or cDNA (complementary DNA). cDNA is enzymatically synthesised from mRNA (messenger RNA) using reverse transcriptase (Russel 1990). The construction of genomic DNA or cDNA libraries involves the cloning of DNA fragments. For genomic DNA libraries, 500 bp to 3 kbp of enzymatically digested DNA fragments are selected and cloned into plasmid vectors. Transformed clones are selected by particular antibiotic(s) (antibiotic selection). IPTG and X-gal, the two key chemicals in the medium, facilitate the differentiation of transformants with the insert DNA from non-recombinant clones. Clones with the insert DNA appear as white colonies while the clones without the insert DNA appear as blue colonies. Recombinant clones are picked and multiplied in a suitable media. Before use in RFLP analysis, the insert DNA is amplified and purified using forward and reverse primers followed by purification (Laurie *et al.* 1997).

The selection of suitable enzyme is an important feature in the construction of a library. The first step in the RFLP analysis is to derive a set of clones that can be used to identify RFLPs. Genomic clones that

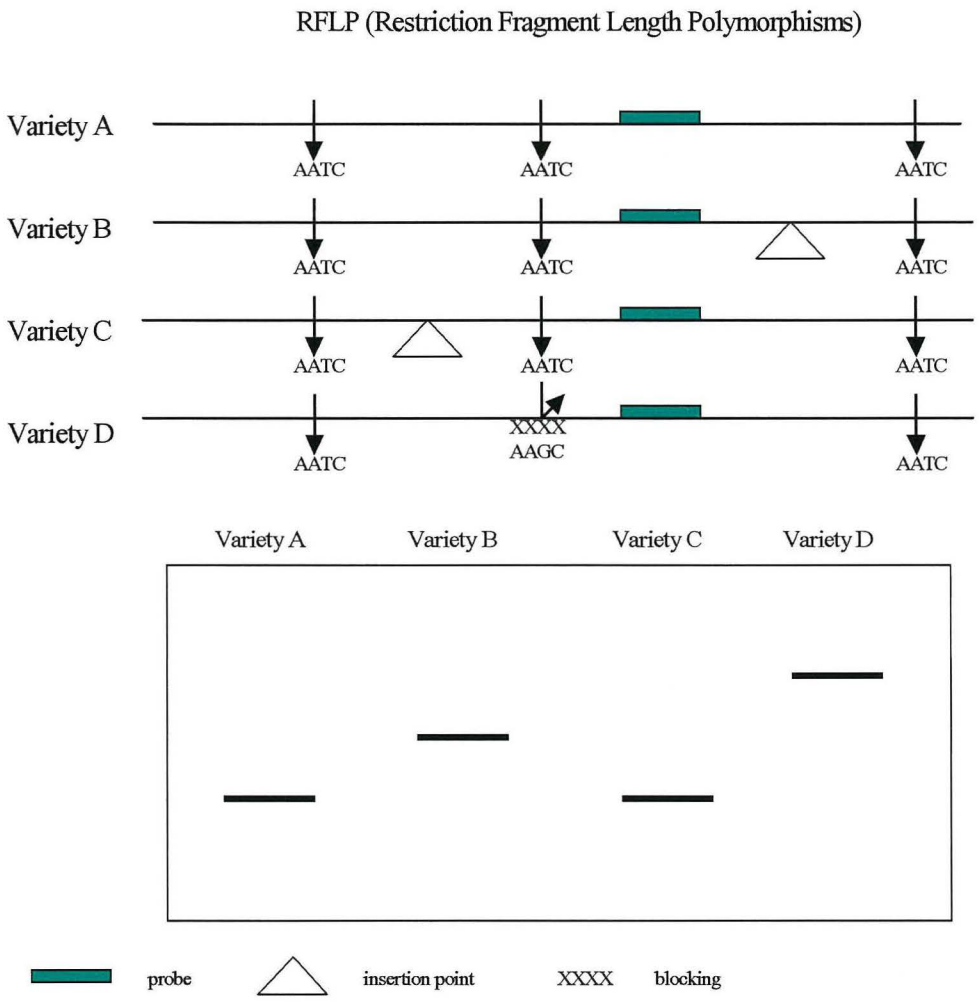


Figure 1.1: Detection of DNA markers by the method of restriction fragment length polymorphism (RFLP). DNA is extracted from four different individuals, and is digested by a restriction enzyme, which cleaves the DNA in or near a specific recognition sequence. In the above example, a restriction enzyme recognises its cleavage site (AATC) on the DNA of four varieties. Variety A has no mutations and represents the basic form of the fragment. Variety B and Variety C have an insertion, within a probed sequence and within a non-probed sequence, respectively. Variety D differs in DNA sequence at one potential recognition sequence. The restriction enzyme cuts in all varieties except D, and generates restriction fragments of different lengths. The fragments are separated by gel electrophoresis, and visualised by the binding of a specific radioactive DNA probe. In practice, RFLP markers are found empirically, by randomly testing different DNA probes with different restriction enzymes until a combination is found which distinguishes between the genotypes of interest (Paterson *et al.* 1991).

represent sequences at random are not suitable as hybridisation probes because plant genomes consist of a large percentage of repeated sequences. Thus, many of the genomic clones contain repeated sequences and these interspersed sequences generate many hybridisation bands that are difficult to analyse genetically. In some cases, hybridisation with these larger clones detects a couple of polymorphic bands from these repetitive sequences. However, a different subset of fragments may reveal as polymorphic when the map is transferred to a different population. The location of these fragments will be unknown. Therefore, low copy sequences are desirable to construct maps. They also can be applied to other populations without much difficulty.

Primary sources of clones for RFLP mapping of plants are cDNA clones and *Pst*I derived genomic clones. The sources of these two clones generally represent expressed genes which are in low copy number. cDNA clones are DNA copies of expressed genes and *Pst*I clones are based on the suggestion that expressed genes are not methylated.

*Pst*I restriction enzyme, which targets undermethylated regions of the genome, has been demonstrated to generate a large proportion of low copy sequences in many crop species (Burr *et al.* 1988). In plants, GC and GXC methylation is the most prominent form of methylation. The enzyme *Pst*I is C-methylation sensitive and will not cleave the DNA if the C at the 5' end of *Pst*I recognition sequence (5'– CTGCAG – 3') is methylated. Therefore, the larger fragments of DNA cleaved using *Pst*I are more likely to contain repeated sequences. These large fragments are excluded from the library by choosing only smaller (300 bp- 1500 bp) fragments of digested DNA, thus resulting in a library for enriched low copy sequences. The remaining clones containing repeated sequences can be removed by screening with total genomic DNA as probes. Clones with repetitive sequences give a strong signal and can be discarded (Liu *et al.* 1994).

RFLP has been successfully used to construct linkage maps of various crop species. RFLP are used extensively in diagnostics, mapping, verifying interspecific hybridisation, and in studying genetic relationships and structure. Probes that hybridise across species are especially useful for comparative mapping (Lefebvre and Chèvre 1995). However, they present some drawbacks that limit their use, including the large amount of DNA that is required for southern hybridisation, the use of radioisotopes, the labouriousness of the technique, and the difficulty in applying them to species with low levels of polymorphism.

- Variable Number of Tandem Repeats (VNTR)

Restriction site analysis of repetitive DNA, VNTR assays, are widely used, especially in forensics. The repeat units are usually less than 100 nucleotides long, with tens to hundreds of copies per locus. Thousands of loci in a genome may have similar core repeat units. The number of repeat units at a VNTR locus can vary greatly between individuals and populations. In plants, the number of repeat units per locus is less variable than in animals. Plant VNTRs are useful markers for variation between and within species. Target fragments are detected by probing with a known repeat sequence (Lefebvre and Chèvre 1995).

B. Polymerase chain reaction (PCR)-based markers

The polymerase chain reaction (PCR) technique was developed by Mullis and Faloona (1987). Compared with RFLP analysis, PCR analysis is easy and rapid, and requires little DNA. PCR involves the principles of DNA reassociation and the action of a thermostable DNA polymerase to amplify the nucleic acid fragment *in vitro.*

(i) Types of PCR-based markers

The principle of the PCR technique gave scope for development of numerous PCR-based marker assays. PCR-based polymorphisms can be random or specific. They can be categorised by whether target sequences are known prior to amplification, whether the primer sequences are designed or are arbitrary, by the number of primers, the size range of amplified products, and by the method of fragment preparation and detection. The most prominent techniques, which have potential for variety identification and genome mapping, include sequence-tagged-sites (STS), microsatellites or simple sequence repeats (SSR), randomly amplified polymorphic DNA (RAPD), and amplified fragment length polymorphisms (AFLP).

- Sequence Tagged Sites (STS)

Oslon *et al.* (1989) defines an STS as a PCR-amplified unique sequence that identifies a known location on a chromosome. A pair of primers is used in each STS reaction and the primer lengths are normally 18-24 bp long. These primers have specific sequences constructed from either previously published DNA sequences or from sequences of anonymous DNA fragments. The majority of STS detect DNA variation at a single locus, that is, they generate a single band from each haploid genome that contains one copy of all genetic material of a given species. Liu *et al.* (1996) and Money *et al.* (1994) indicate that direct STS analysis often reveals only a limited level of variations, which results in narrowing its general usefulness in variety identification. However, it may be the technique of choice in identifying varieties bred by adding a known piece of DNA sequence (gene) to existing varieties through transformation. Furthermore, STS expresses co-dominance thereby allowing the genotype at any locus to be determined in any breeding scheme.

- Simple Sequence Repeats (SSR) or microsatellites

SSR are tandem repetitive DNA sequences of usually di- or tri- nucleotide repeats flanked by unique sequence DNA (Laurie *et al.* 1997). They were first referred to as microsatellites by Litt and Luty (1989) and later as single sequence repeats (SSRs) by Jacobs *et al.* (1991). Primers, complementary to unique sequences flanking the microsatellite, are used to amplify the microsatellite. The polymorphism revealed is due to the change in the number of repeats. For example, an $(AT)_5$ repeat (ATATATATAT) is a microsatellite resulting from the repetition of two nucleotides five times. In one variety the two nucleotides could repeat ten times, $(AT)_{10}$, and in another variety, the two nucleotides could repeat 20 times, $(AT)_{20}$. The procedure used for SSR analysis is the same as STS analysis, that is, it needs two specific primers for each reaction. The only difference between the two marker systems is that the former targets highly polymorphic regions of a genome, whereas the latter is nonselective for amplified random sequences with respect to the level of polymorphism. SSR have been generally recognised to be an excellent marker system. These markers reveal a higher incidence of detectable polymorphism and are more informative than any other DNA marker (Paglia *et al.* 1998). In addition, they have been abundant, have a uniform distribution throughout the genome (Wang *et al.* 1994), and reveal co-dominantly inherited multi-allelic products of loci that can be readily mapped. This creates an advantage over most PCR methods that are based upon the amplification of arbitrary sequences that usually originate from diverse loci (Cordeiro *et al.* 2001)

- Randomly amplified polymorphic DNA (RAPD)

RAPD was first described by Welsh and McClelland (1990), and Williams *et al.* (1990). Unlike STS and SSR marker systems, RAPD analysis uses a single type of arbitrary primer, 9-10 nucleotides long. This primer anneals to the template at complementary sequences in both '+' and '-' strand

orientation allowing the amplification of several bands corresponding to several loci. Some unique advantages of RAPD marker system include (i) data is not needed for synthesising primers, and (ii) many primers can be used for PCR amplification of DNA from a wide range of species. Also, several markers can be detected from a single RAPD reaction. Disadvantages with RAPD markers are that, for the presence of a fragment, homozygous dominant individuals cannot be distinguished from heterozygous dominant, and the progeny of backcrosses with the dominant parent does not segregate (Lefebvre and Chèvre 1995). It has also been reported that results of RAPD amplification from the same primers can vary, not only between different laboratories but also between different thermocyclers. Further, the system is prone to a high degree of error due to template competition (Penner *et al.* 1993).

- Amplified fragment length polymorphisms (AFLP)

AFLP combines aspects of RFLP and RAPD (Vos *et al.* 1995). There are three main steps in the AFLP procedure. First, genomic DNA is digested with two different restriction enzymes followed by ligation of adaptors to the ends of the fragments. These digested DNA fragments, with adaptors of known sequences, are then used as a template for a PCR reaction. They are amplified with primers containing common sequences of adaptors, but with one to three additional nucleotides (called selective nucleotides). This ensures that only a fraction of the restriction fragments are amplified. Amplified fragments are then analysed on a polyacrylamide gel. Advantages of AFLP analysis are that a single reaction can detect a large number of loci and that the number of loci detected per reaction can be regulated by varying the number of the selective nucleotides (van Eck *et al.* 1995). Then, like SSR and STS analysis, AFLP analysis uses a pair of specific primers resulting in more reliable and reproducible results than RAPD. However, AFLP is the most complicated PCR-based marker system, involving DNA restriction digest and ligation reactions, thus

requiring high quality DNA. Further, AFLP markers are dominant markers like RAPD, hence less informative than many other markers types (Zabeau and Vos 1994).

(ii) Other PCR- based marker systems

- Sequence characterised amplified regions (SCAR)

SCAR (Paran and Michelmore 1993) is considered the second generation of RAPD markers. Primers are synthesised based on the sequence of the ends of RAPD fragments to generate specific PCR markers. Polymorphism is directly detected in the case of length polymorphism or after cleaving with a restriction enzyme (Lefebvre and Chèvre 1995). The usefulness of using SCAR has been shown in lettuce, where primers were designed from nine RAPD markers found to be linked to downy mildew resistance genes (Paran and Michelmore 1993)

- PCR-SSCP (single-stranded conformation polymorphism)

PCR-SSCP involves amplification of the target sequence with radioactively labelled primers or nucleotides. The amplified product is denatured to a single-stranded form prior to electrophoresis on a non-denaturing polyacrylamide gel (Orita *et al.* 1989). In non-denaturing conditions, single stranded DNA has a folded structure, which is determined by the nucleotide sequence (Kanazawa *et al.* 1986). Any mutations result in bands of single-stranded DNA at different positions. PCR-SSCP is advantageous to detect polymorphisms in regions of interspersed repetitive sequences, which are more amenable to mutations.

- Arbitrary primed PCR (AP-PCR)

AP-PCR uses primers with an 18-24 bp length, and the amplification products are detected on agarose gels after staining with ethidium bromide (Welsh and McClelland 1990).

- DNA amplification fingerprinting (DAF)

DAF uses very short primers, usually 8-mers, but some as short as 5-mers can be also used. Amplification products are separated on a polyesterbacked polyacrylamide gel containing 7 M urea and are detected by silver staining, resulting in a 2- to 3- fold increase in the number of polymorphic and monomorphic fragments (Caetano-Anolles *et al.* 1991).

- Inter-simple sequence repeat amplification (ISSR)

The ISSR marker system amplifies the DNA segments that lie between two opposed SSRs. PCR amplification of the target region is carried out with a terminally anchored primer of the same type, which is in part complementary to the flanking microsatellites (Zietkiewcz *et al.* 1994). Polymorphism results from whenever one genome is missing one of the SSR or has a deletion or insertion that modifies the distance between the repeats (Salimath *et al.* 1995). The radiolabelled amplification products are visualised by autoradiography. The ISSR system is able to develop a large number of polymorphic markers. However the dominant nature of the marker makes it less informative.

- Specific polymorphic locus amplification test (SPLAT)

Primers for SPLAT are designed from sequencing the ends of RFLP polymorphic fragments, hence it is considered to be a second generation RFLP marker system (Gale and Witcombe 1992).

- Single-nucleotide polymorphisms (SNP)

SNPs are defined as single-base variation in a DNA sequence, usually represented as two or sometimes three different bases at a single position (Nowatny *et al.* 2001). SNPs have been used in human genetic analysis, including forensic analysis, comparative genetics and evolution studies. The development and application of SNPs as genetic markers in plant have been rapidly increasing (Rafalski 2002). SNPs are highly abundant,

stable (i.e. less prone to the 'slippage' seen with microsatellite repeats, Nowatny *et al.* 2001). In addition, many assays have been developed to type SNPs in an automated fashion and many yield simple positive or negative results that can be interpreted easily (Kota *et al.* 2001).

1.2.3.3 Ideal Markers

The selection of appropriate marker assays can be challenging. Until the discovery of genetic markers at the protein level in the 1960s, morphological markers, which produced clearly visible effects on the plant's phenotype, were the only genetic markers used. As described previously, morphological markers have many disadvantages. They often affect the fitness of the individual and exhibit pleiotropy, mitigating their value as a genetic tool.

Discovery of naturally occurring genetic polymorphisms at the protein level has made genetic studies easier. Variation in enzymes and storage proteins is detectable on an electrophoresis gel for a very high proportion of gene loci. Such naturally occurring allelic variants have much smaller effects on fitness. However, the number of polymorphisms at the protein level is limited and can not compete with PCR-based molecular marker systems.

The vast amount of the genetic polymorphism which exists at the DNA level, revealed after the development of the PCR technique, has revolutionised genetic analysis. As mentioned before, such variation results from the existence of occasional base changes in the DNA, which can be recognised by restriction enzymes or primers used in PCR analysis. Depending on how the DNA polymorphism is studied, various types of markers are used. For example, SSR or AFLP marker systems are the most suitable for variety identification. For mapping studies,

polymorphism in repeated sequences (SSR or microsatellites) is useful. However, these marker types are often used in combination with other systems such as isozymes or DNA fragment electrophoresis, for genetic mapping. Similarly, a combination of chromosome banding, ISH, and STS markers is used for physical mapping (Westman and Kresovich 1997).

Although PCR-based markers can be used for mapping, most reasonably detailed maps are based on analysis of restriction fragment polymorphism (RFLP), which is detected by the hybridisation of cloned DNA segments to genomic DNA digested with restriction enzymes. They have the advantage of being codominant, enabling all genotypes in a cross to be identified. However, as with other marker systems, RFLP analysis has some drawbacks. RFLP analysis requires a large quantity of DNA, and involves the use of radioactive probes, which incurs safety considerations. In addition to these considerations, it is a time consuming process.

Therefore, no marker system can be ideal for every situation. The selection of appropriate markers depends on the purpose of the studies, as well as existing laboratory set-ups, skills and viewpoints.

1.3 Linkage Map and Mapping Populations

A linkage map is not essential for marker application. However a map-based approach can dramatically enhance the reliability of a project because it is the only way to find out if markers covering a whole genome have been utilised. A linkage map also makes it possible to utilise markers systematically for a given species. Cross-checking results from different experiments makes a marker project more efficient. This is the reason why, if resources permit, linkage map construction is often the first phase of a mapping project. Table 1.2 shows a list of linkage maps for some major crop species.

Table 1.2: Molecular linkage maps of major crops

Crop species	Reference
Rice	Harushima *et al.* 1998
Hexaploid wheat	Nelson *et al.* 1995
Triticum tauschii	Gill *et al.* 1991
Oats	O'Donoughue *et al.* 1995
Barley	Liu *et al.* 1996
Rye	Loarce *et al.* 1996
Pearl millet	Liu *et al.* 1994
Maize	Beavis and Grant 1991
Sorghum	Xu *et al.* 1994
Sugarcane	Mudge *et al.* 1996
Brassica oleraces	Landry *et al.* 1992
B. rapa	Chyi *et al.* 1992
B. campestris	McGrath and Quiros 1991
B. napus	Foisset *et al.* 1996
B. juncea	Cheung *et al.* 1997
Arabidopsis thaliana	Chang *et al.* 1988
Tomato	Grandillo and Tanksley 1996
Potato	Tanksley *et al.* 1992
Soybean	Keim *et al.* 1997
Mungbean	Menancio-Hautea *et al.* 1993b
Cowpea	Menedez *et al.* 1997
Common bean	Vallejos *et al.* 1992
Pea	Gilpin *et al.* 1997
Chickpea	Ratnaparkhe *et al.* 1998
Lentil	Weeden *et al.* 1992
Lettuce	Kesseli *et al.* 1994
Cassava	Fregene *et al.* 1997
Cotton	Shappley *et al.* 1996
Coffee	Paillard *et al.* 1996
Cucumis melo	Baudracco-Arnas and Pitrat 1996
Papaya	Sondur *et al.* 1996
Sugar beet	Barzen *et al.* 1995

1.3.1 Basic Principle of Linkage Map Construction

A genetic linkage map is a graphical representation of an array of loci, which may include morphological and isozyme markers as well as DNA markers, along the chromosome. It is developed following the analysis of a large number of markers in segregating progeny of polymorphic parents. Recombination frequencies between pairs of markers can be scored for each of the individuals and the recombination values for the population are used to estimate the distance between these markers (Kearsey 1997). The distance between these markers is expressed in centimorgans (cM), which represents the recombination rates between the markers (1 cM ≈ 1% recombination). This estimated distance for linkage map construction has no precise relationship with the physical distance expressed in base pairs (bp) because the rate of recombination can vary along the length of chromosome (Kumer 1999; Lefebvre and Chèvre 1995).

Two types of possible scoring methods exist for any markers – either present/absent or A/B/H (Figure 1.2). If marker 1 and 2 give identical score in a given population, the two markers are said to map to the same position. If the score of a third marker (3) has no correlation with these two markers (i.e. >50% recombination frequency, Figure 1.3a), it maps to a different chromosome or to a chromosome region that recombines freely with the section occupied by the first two markers (Figure 1.3a). Of course, the above are two extremes. In between, it is possible to get markers that do not map together but are linked together with a degree of recombination. For example, if marker 1 and marker 2 have a recombination frequency of 6%, they will map 6 cM apart. Then, if marker 3 has a recombination frequency with marker 1 of 10%, it will map 10 cM from marker 1. It also has a recombination frequency of 15% with marker 2, then it will map 15 cM from marker 2. Knowing the recombination frequency of each marker with each other, they can be ordered along the

Case 1: present/absent score

P_1	P_2	population
		1 2 3 4 x
-	x	- - x -.....-
C	A	C C A C .....A

Case 2: A/B/H score

P_1	P_2	population
		1 2 3 4.....x
-	-	- - - -.....-
A	B	A H B B..... H

Figure 1.2: Two ways of scoring markers in a population. Case 1 is a dominant marker system, scored by the presence or absence of a band. Data is put into software as "C" for presence of a band and "A" for absence of a band. Case 2, A/B/H score represents expression of a codominant marker in a population. As "A" for one of parental expression and "B" for the another. Heterozygous in a population is transcribed as "H".

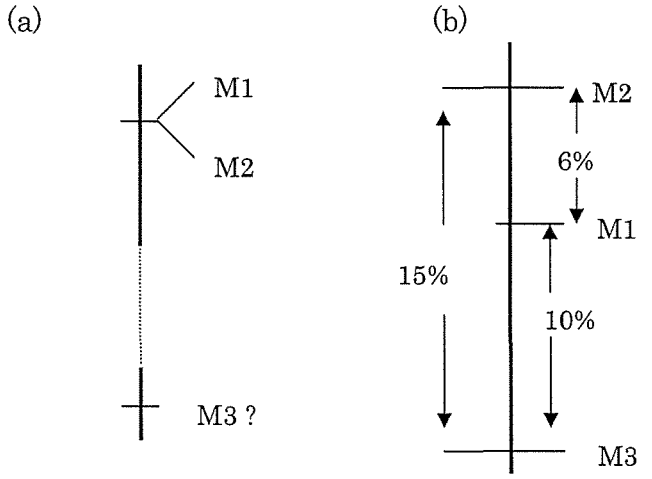


Figure 1.3: Basic principle of map construction. (a) If markers 1 and 2 give identical score in a given population, they are mapped at the same position. If a third marker (M3) has no correlation with these two markers, it maps to a different chromosome or to a chromosome region which recombines freely with the section occupied by the first two markers. (b) If markers 1,2 and 3 have recombination frequencies of between 0% and 50%, their relative positions are mapped accordingly.

chromosome (Jones *et al.* 1997). Figure 1.3 (b) explains this. Due to double recombination events, some recombination frequencies are not additive (e.g. marker 2 and marker 3). When a large number of markers are tested against each other in a mapping population, a map can be constructed from their relative positions. If a map in which the number of linkage groups equals the number of chromosome pairs and there are no unlinked markers, it is often referred to as a "perfect map". However, apart from a few well studied species, most linkage maps produced so far are not perfect. The numbers of linkage groups in most maps are larger than the number of chromosome pairs and they often contain unlinked markers. Those markers on a given chromosome appear as two or more separate linkage groups simply because the subsets are not sufficiently close for them to be recognised as being together on one chromosome (Kearsey 1997).

A range of software packages such as MAPMAKER (Lander *et al.* 1987) and JOINMAP (Stam 1993) have facilitated the construction of linkage maps. These software packages estimate the recombination frequencies, identify linkage groups, assign the markers to the most likely order, and then space them in map units (cM).

1.3.2 Segregating Populations Most Commonly Used for Map Construction

Recombination values between pairs of markers can be scored using a wide variety of different cross types. The populations that are most commonly used for map construction are F2, recombinant inbred lines (RIL), and doubled haploid (DH) lines.

1.3.2.1 F2 Population

An F2 population is produced by firstly crossing two selected parents to generate a single F1 hybrid, and then self-crossing the F1 hybrid to get F2 individuals. F2 populations are easy to obtain (apart from those species that are self-incompatible) and have been widely used for generating maps for many species. However, as not all loci in an F2 are fixed this population cannot be used in replicated experiments that are often essential in tagging genes conditioning quantitative traits. This inherent problem of F2 populations can be partially overcome by utilising bulked F3 families (Soller and Beckmann 1990).

1.3.2.2 Recombinant Inbred Lines (RIL)

RIL populations are produced by selfing F2 individuals for many (normally eight or more) generations. By then, the majority of loci would be fixed for each individual. Sometimes it may take several years to produce a RIL population depending on the life cycle of the crop. However, once produced, RIL populations are very valuable, not only in genome mapping but also in gene tagging. This is because a large number of seeds with similar genetic composition can be produced from each individual, and this can meet the requirements of using replicated trials under different locations, and at different seasons for those traits conditioned by many loci (Kearsey 1997).

1.3.2.3 Doubled Haploid (DH) Population

DH populations can be generated in different ways. Another culture and wide hybridisation are two widely used approaches. For example, the hybrid embryos produced when hexaploid wheat is pollinated with maize are karyotypically highly unstable and lose all maize chromosomes in the

first few cell division cycles. The haploid embryos produced can be recovered using spikelet culture and fertile DH plants are recovered after colchicine treatment (Laurie 1988). Similar to RIL, all loci for DH lines are fixed and this type of population is suitable for investigating the genetics of complex traits. However, unlike RIL, DH lines can be generated in a short time. However, the tissue culture and colchicine treatments involved in producing DH lines can be technically demanding for some species.

1.4 Marker and Linkage Map Applications

1.4.1 Diversity and Phylogenetic Studies

Diversity and phylogenetic investigations have long been important components in biology. Results from these studies provide the scientific basis for efficient utilisation/conservation of extant genetic variation (Schoen and Brown 1993). Before the invention of genetic markers, diversity and phylogenetic studies were based on morphological differences, e.g. flower colour and plant height. With the introduction of molecular markers it has become possible to detect individual alleles at an individual locus, thus the ability to reveal detailed variation and phylogenetic relationships between taxa has been dramatically enhanced. With the application of molecular markers in these fields, we have learnt a great deal about diversity in a wide range of species (e.g. Liu 1996) and our understanding of the evolutionary relationships between many taxa have been dramatically improved (e.g. Liu and Musial 1997). Results from these studies have also provided some new strategies in more efficiently utilising some taxa in plant improvement (Liu *et al.* 1999; Liu and Musial 2001).

1.4.2 Gene Tagging/QTL Mapping

A direct application of genetic linkage maps has been in tagging genes of economic importance with molecular markers. Markers for genes of agronomic importance are powerful tools in breeding programs (Tanksley 1993). Before the introduction of molecular markers, other marker types had been widely used for marker development. However, due to the limitation of available markers, success was limited. One of the few successful cases was the identification of a protein marker for eye spot resistance in wheat, which has been widely used in European wheat breeding programs (Summers *et al.* 1988). In general, the likelihood of identifying a marker linked to a gene is inversely proportional to the distance between the marker and the gene (Kumer 1999). As mentioned previoualy, the use of software packages, which compute and analyse variance models, can perform the detection of linkages between markers and genes. However, quantitative traits such as yield, quality, height, maturity, and resistance to several biotic and abiotic loci are controlled by several genes, each of which makes a small positive or negative contribution to the final phenotypic value of the trait. In addition, such traits are greatly influenced by the environment. Genes underlying such traits are called either polygenes, effect factors, or, more recently, quantitative trait loci (or QTL; Kearsey 1997). With the introduction of molecular markers, there have been extensive efforts to develop markers for different traits of importance in many species (e.g. Chao *et al.* 1989). It is now possible to assign chromosome locations to individual QTLs, and also to determine which parent possesses the positive alleles at each QTL (Edwards *et al.* 1987). In addition to the direct application of molecular markers in breeding programs, in the process of developing markers it is now possible to learn a great deal about the genetics of a trait (e.g. Thumma *et al.* 2001). This in turn can provide scientists with ideas of how to manipulate these traits more efficiently.

1.4.3 Comparative Mapping

One of the most useful applications of genetic linkage maps has been in the comparison of genomes of distantly related or cross incompatible taxa. Comparative mapping has become an important branch of modern genomics. It offers the possibility to compare genome structures between productively isolated species. Using this approach, genome structures among many plant species have been compared. For example, it has become clear that genomes of wheat, rye and barley are highly conserved, and they also share a very high level of synteny with other grass species like rice, maize, sorghum and millet (Gale and Devos 1998a). These studies not only dramatically improved our understanding of the mechanisms involved in the evolution of these species, but they also make it possible to predict the location of genes of importance. This knowledge has been widely used for gene tagging (e.g. Asnaghi *et al.* 2000).

1.4.4 Optimising Marker Application

Linkage map construction is time-consuming and expensive. So, only one or a small number of good linkage maps for a given species can be constructed (in fact, for the reasons discussed below, there is no need to construct more than one good map for any given species). However, the number of traits that can be studied in any given population is limited, while the number of genotypes exploited by breeders and many other scientists can reach tens of hundreds or even more. This requires markers and their mapping information to be transferred constantly to new populations. It is important that markers covering a whole genome are exploited in a new mapping project. It is believed that some 10 to 20 cM between markers would be adequate for gene tagging given the practical population sizes currently used (Tanksley 1993). In other words, it is the marker distribution, not the number of markers that is important in a

mapping project. So, ideally a new marker project should be carried out by: (1) transferring a small number of evenly distributed markers from an existing map to a new population; (2) using these markers to identify the chromosome region(s) where genes of interest are located; and (3) developing markers for these targeted regions. Figure 1.4 shows the ideal ways of applying map information in gene tagging.

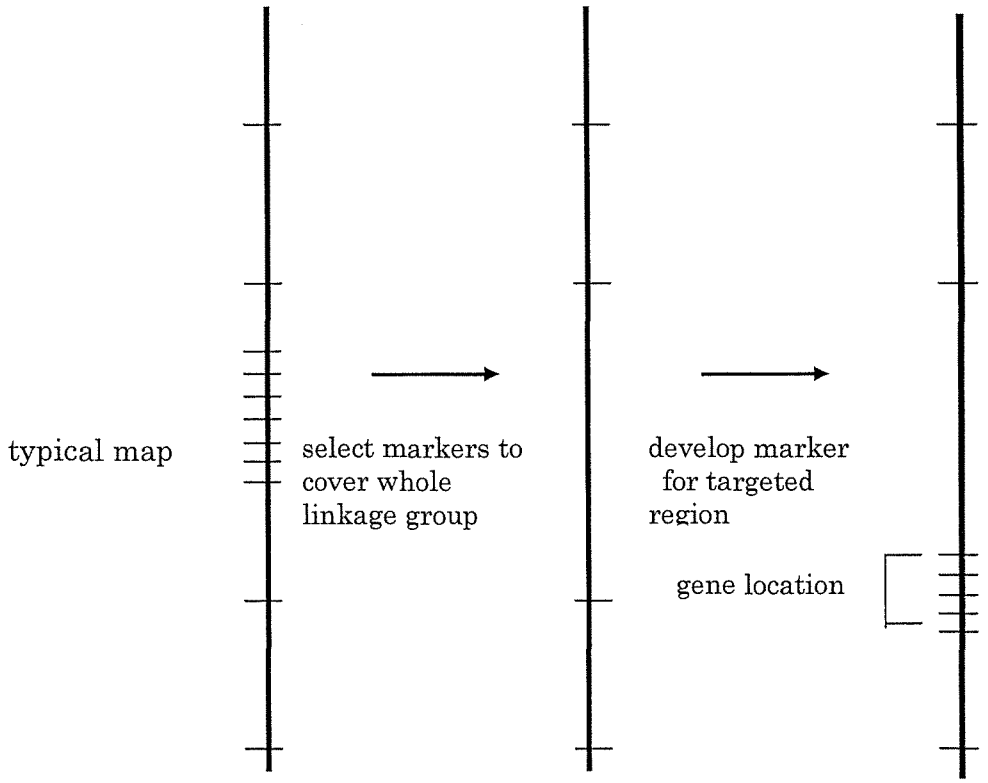


Figure 1.4: Ideal ways of applying map information in gene tagging. It involves three steps including: (1) constructing a comprehensive map, (2) transforming a framework map to a new population and identifying region where gene of interest locates, and (3) developing markers closely linked to the gene of interest.

1.5 Bacterial Artificial Chromosome (BAC) and Other Large Insert Cloning Systems

Large insert genomic DNA libraries, using vectors such as cosmids (Collins and Hohn 1978), yeast artificial chromosomes (YACs) (Burke *et al.* 1987), bacterial artificial chromosomes (BACs) (Shizuya *et al.* 1992), and P1-derived artificial chromosomes (PACs) (Iannou *et al.* 1994), have played a pivotal role for the isolation and characterisation of important genomic regions and genes (Zhang *et al.* 1996b). Based on the hosts used for these large-insert DNA clones, they can be classified as bacteria- and yeast-based cloning systems. YAC belongs to the yeast-based cloning system, while cosmid, BAC and PAC belong to the bacteria-based cloning system (Zhang and Wu 2001).

1.5.1 Cosmids

Cosmids were the first large-insert DNA cloning system invented (Collins and Hohn 1978). A cosmid is a plasmid cloning vector utilised for isolating a large and complex genomic DNA. This vector contains a bacteriophage lambda cos site that directs insertion of DNA into phage particles (Monaco and Larin 1994). Cosmids allow the cloning and maintenance of DNA fragments of about 40kbp in bacteria, using the conventional bacteriophage-based transfection method to deliver its constructs into bacterial cells. They were once widely used in genome-wide physical mapping, positional cloning and large-scale sequencing of large and complex genomes.

1.5.2 YAC

Since the YAC system was first reported by Burke *et al.* (1987), YACs have revolutionised genome research. YACs are linear constructs and each YAC

clone consists of all features of a typical chromosome, including two chromosomal arms, a centromere and two telomeres. This has allowed the cloning and maintenance of DNA fragments over 1000 kbp (Zhang and Wu 2001). YAC libraries have been constructed for a number of plant species including *Arabidopsis* (Ecker 1990; Grill and Somerville 1991; Ward and Jen 1990), maize (Edwards *et al.* 1992), tomato (Martin *et al.* 1992), and rice (Umehara *et al.* 1995). These YAC libraries have been extensively used for physical mapping (Zhang and Wing 1997) and positional cloning (Arondel *et al.* 1992; Martin *et al.* 1993). However, there are two major limitations in utilising YAC libraries. The first one is that YAC libraries have been found to contain significant amounts of chimeric clones (Anderson 1993). These are clones in which a part of the insert comes from one region of the genome while another part comes from a different region. Thus, it is impossible to assign these clones into specific chromosome regions, which makes them difficult to use. The second problem with the YAC system is that it is not always easy to separate an insert from yeast chromosomes. Yeast has sixteen chromosomes ranging from 230 to 1532 kbp in size, YACs that have insert sizes falling within the yeast chromosome size range are difficult to separate from the yeast chromosomes (Zhang and Wu 2001). It may take 3-5 days to isolate a YAC insert and DNA yields are often very low (Zhang *et al.* 1996b).

1.5.3 BAC and PAC

Shizuya *et al.* (1992) reported large DNA fragment cloning in *Escherichia coli* using a bacterial artificial chromosome (BAC) system based on the *E. coli* fertility (F-factor) plasmid. Two years later, Iannou *et al.* (1994) developed the PAC system, which combines the features of the bacteriophage P1 and the F-factor-based BAC cloning system. Compared to YAC, insert sizes of BACs and PACs are somewhat smaller. Both BAC and PAC are capable of cloning DNA fragments of up to 400 kbp in *E.coli*

(Zhang *et al.* 1996b). However, these systems are a lot more "user-friendly" than YAC. BAC and PAC have a low level of chimerism, and isolation and purification of insert DNA from these clones is very easy (Woo *et al.* 1995). Also, high stability of inserts in these host cells has been reported (Table 1.3 compares YAC and BAC). Two widely used vectors are pBeloBAC11 (Figure1.5) and its derivative pECBAC1 (Zhang and Tao 1998)

Table 1.3: Comparison between YAC and BAC cloning systems (http://www.genome.clemson.edu).

Features	YAC	BAC
Configuration	Linear	Circular
Host	Yeast	Bacteria
Copy Number/ Cell	1	1-2
Cloning Capacity	Unlimited	Up to 400 kbp
Transformation	Spheroplast (10^7 T/µg)	Electroporation (10^{10} T/µg)
Chimerism	Up to 40%	None to low
DNA Isolation	PFGE Gel Isolation	Standard Plasmid Miniprep
Insert Stability	Unstable	Stable

The basic structure of the BAC/PAC vectors is derived from the endogenous plasmid F. The F backbone contains four essential regions that function in plasmid stability and copy number. The PAC vector has most of the features of the BAC system, however the vector contains the SacB gene, which facilitates a positive selection for recombinant clones during library construction. SacB encodes sucrose synthase. Therefore when cells are grown in the presence of saccharose, sucrose synthase will degrade saccharose into levan, which is highly toxic to *E.coli*. The *Bam*HI

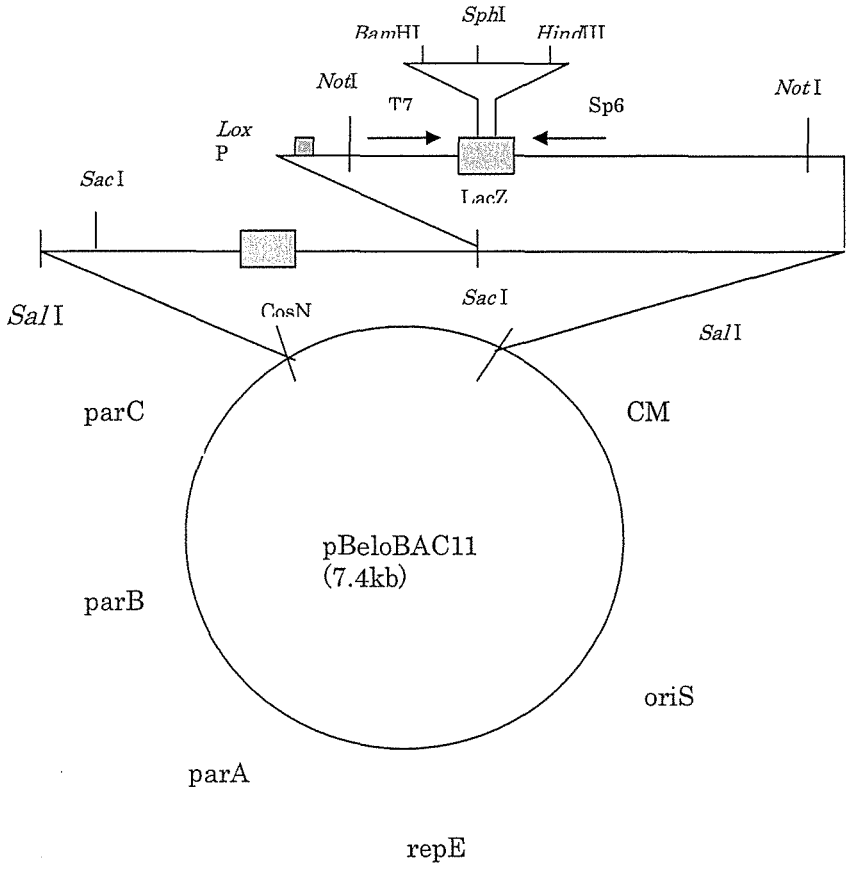


Figure 1.5: Diagram of pBeloBAC11 vector. The plasmid is based on a mini-F plasmid, pMBO131. Both ParA and ParB are required for partitioning and plasmid stability. ParB is also required for incompatibility with other F factors. OriS is the origin of DNA replication, which is unidirectional. RepE encodes protein E which is essential for replication from OriS and for copy number control. A chloramphenicol resistance gene was incorporated for antibiotic selection of transformants. The identification of recombinant DNA clones is simplified with lacZ gene for colour selection. CM^r = chloramphenicol resistance. (http://www.tamu.edu)

cloning site is within the SacB gene, resulting in disruption of the SacB gene by insertion of a large DNA fragment. This allows growth of the cells on media containing sacchrose. Moreover, the vector carries a "PUC19-link", containing a high copy number origin of DNA replication, which is used for convenient vector propagation and is later removed during vector preparation for library construction.

BAC has become the technique of choice for large DNA clones during the last few years. BAC libraries have been constructed for many species and these libraries have been extensively used for a wide array of endeavours, including physical mapping (Zhang and Wing 1997), map-based gene cloning (Arondel *et al.* 1992; Bent *et al.* 1994; Martin *et al.* 1993), as well as in developing molecular markers (Cregan *et al.* 1999; Martin *et al.* 1992; Zhang H.B. *et al.* 1994). For a given genome, the large insert sizes of the source clones reduce the number of clones needed for mapping. To a great degree, the genome coverage of a map depends on the extent of the genome representation of the source library. The genome representation of a library is calculated with the number of clones, times the average insert size, divided by the genome size.

However, Zhang and Wu (2001) reported that the theoretical genome coverage of a large-insert BAC library is often lower than its true genome coverage by about 15%, and the true genome coverage of the library cannot be significantly increased by simply increasing the number of clones in the library. Table 1.4 shows the number of clones that are needed to have a probability of 99% that a library contains a desired clone, based on average insert size of 40, 150 and 500 kbp for cosmid, BAC or YAC library, respectively, for some crop species.

Table 1.4: The number of clones required for a 99% probability that a particular clone is represented in a library having an average insert size of 40kb, 150kb or 500 kb for selected crop plants (Choi and Wing 2000; Zhang *et al.* 1996b).

Scientific Name	Common Name	Genome Size Mbp/1C	Cosmids 40 kbp	BACs 150 kbp	YACs 500 kbp
Allium cepa	Onion	15,290	$1.8x10^6$	$4.7x10^5$	$1.4x10^5$
Arabidopsis thaliana	Arabidopsis	145	$1.7x10^4$	$4.5x10^3$	$1.3x10^3$
Avena sativa	Oat	11,315	$1.3x10^6$	$3.5x10^5$	$1.0x10^5$
Beta vulgaris ssp. *esculenta*	Sugar beet	758	$8.7x10^4$	$2.3x10^4$	$7.0x10^3$
Brassica napus	Canola	1,182	$1.4x10^5$	$3.2x10^4$	$1.1x10^4$
Capsicum annuum	Pepper	2,702	$3.5x10^5$	$9.4x10^4$	$3.5x10^3$
Citrus sinensis	Orange	382	$4.4x10^4$	$1.2x10^4$	$3.5x10^3$
Glycine max	Soybean	1,115	$1.3x10^5$	$3.4x10^4$	$1.0x10^4$
Gossypium hirsutum	Cotton	2,246	$1.8x10^5$	$6.9x10^4$	$2.1x10^4$
Hordeum vulgare	Barley	4,873	$5.6x10^5$	$1.5x10^5$	$4.5x10^4$
Lactuca sativa	Lettuce	2,639	$3.0x10^5$	$8.1x10^4$	2.4x10
Lycopersicon esculentum	Tomato	953	$1.1x10^5$	$2.9x10^4$	$8.8x10^3$
Malus x domestica	Apple	769	$8.9x10^4$	$2.4x10^4$	$7.1x10^3$
Manihot esculenta	Cassava	760	$8.7x10^4$	$2.3x10^4$	$7.0x10^3$
Musa sp.	Banana	873	$1.0x10^5$	$2.7x10^4$	$8.0x10^3$
Nicotiana tabacum	Tobacco	4,434	$5.1x10^5$	$1.4x10^5$	$4.1x10^4$
Oryza sativa ssp. *indica* & *japonica*	Rice	431	$5.0x10^4$	$1.3x10^4$	$4.0x10^3$
Phaseolus vulgaris	Common bean	637	$7.4x10^4$	$2.0x10^4$	$5.9x10^3$
Saccharum sp.	Sugarcane	3,000	$3.5x10^5$	$9.2x10^4$	$2.8x10^4$
Sorghum bicolor	Sorghum	760	$8.7x10^4$	$2.3x10^4$	$7.0x10^3$
Triticum aestivum	Wheat	15,966	$1.8x10^6$	$4.9x10^5$	$1.5x10^5$
Zea mays	Maize	2504	$2.9x10^5$	$7.7x10^4$	$2.3x10^4$

1.6 General Procedures for BAC Library Construction

The construction of BAC libraries involves:

1.6.1 Megabase DNA Isolation

BAC clones contain inserts up to 400 kbp. To construct such libraries, very high molecular weight (HMW) DNA, "megabase-size", is required. To isolate such DNA, protoplasts or nuclei must first be embedded in agarose plugs or microbeads. The agarose acts as a solid matrix that protects the DNA from shearing. Once embedded, the protoplasts or nuclei are lysed and proteins degraded in the presence of proteinase K. After cell lysis and protein degradation, the remaining DNA is suitable for enzymatic modification.

1.6.2 Generation of Large DNA Fragments from Megabase-Size DNA

This is normally achieved by partial digestion with an appropriate restriction enzyme. After partial digestion, HMW DNA fragments are separated using PFGE (Albertsen *et al.* 1990), and DNA fragments in the desired size range (often from 150 kbp to 400 kbp) are removed from the gel into several sections and stored in EDTA solution at 4°C to prevent DNA shearing.

1.6.3 Vector Preparation

Several BAC vectors have been developed including BAC 11 and True Blue. Most of these BAC vectors are single copy in *E. coli* (Shizuya *et al.* 1992) so a large volume of culture is often needed to get a decent quantity of plasmid. Once isolated, the vector DNA is digested with the same restriction enzyme as used for partial digestion of insert DNA. After

complete digestion, vectors are dephosphorylated to prevent self-ligation.

1.6.4 Ligation to Create Recombinant DNA

After the digestion of insert DNA and cloning vector with the same restriction enzymes, they are ligated together using T4 DNA ligase to generate recombinant DNA molecules.

1.6.5 Transfer of Recombinant DNA Molecules into Host Cells, *E. coli.*

The most widely used *E.coli* strain for BAC cloning is DH10B. Key features of this strain include mutations that block: a) restriction of foreign DNA by endogenous restriction endonucleases (hsdRMS); b) restriction of DNA containing methylated DNA (5' methyl cytosine or methyl adenine residues, and 5' hydroxymethyl cytosine) (mcrA, mcrB, mcrC, and mrr); c) recombination (rec A1) (see details in http://www.tamu.edu). Due to the very large insert sizes of BAC clones, transformation efficiency can be low. Therefore, electroporation rather than the $CaCl_2$ protocol (Hanahan 1983) is routinely used (Shizuya *et al.* 1992).

1.6.6 Clone Picking and Storage

Recombinant clones are identified by colour reaction (white/blue) and libraries are normally stored as cells in 384-well microtitre dishes as glycerol stocks.

1.7 A Strategy for Utilising BAC Technology in Marker Projects

One of the main difficulties in a marker-based project is that there is no

guarantee of being able to transfer a marker from an existing map to a new population because it might not be able to detect polymorphism. This is a common problem in marker projects of any species, even though some genomes (e.g. mungbean, see Humphry *et al.* 2002) are a lot more polymorphic than others, such as wheat (Chao *et al.* 1989), and some marker systems (such as SSR) are more powerful than others (such as RFLP).

It is important to understand that existing marker systems detect a small proportion of DNA variation that covers between a few hundred (all PCR-based markers) to a few thousand (RFLP) base pairs at any given locus. When BAC clones become available, each of these markers can be used as a probe to isolate one or more BAC clones residing in the same region as the initial marker (probe). Because of their huge size (50 to 200 kbp or more), BAC clones can easily be used to generate many RFLP probes or other types of markers, all detecting variation in the same region. If necessary, the ends of the first BAC clone identified can be isolated and used as probes to isolate adjacent BAC clones in the same region (Figure 1.6). Thus, when BAC clones are used, the ability to isolate one or more polymorphic markers for any chromosome region would be dramatically enhanced.

1.8 Conclusion

The development of molecular marker technology has enabled breeders to use a Mendelian genetic approach to complement plant breeding programs. Over the last twenty years, with the advent of DNA marker technology, a variety of DNA markers and molecular marker breeding strategies are now available to plant breeders and geneticists, helping them to overcome many of the problems faced in conventional breeding programs. One major limitation of DNA markers such as RFLP is their

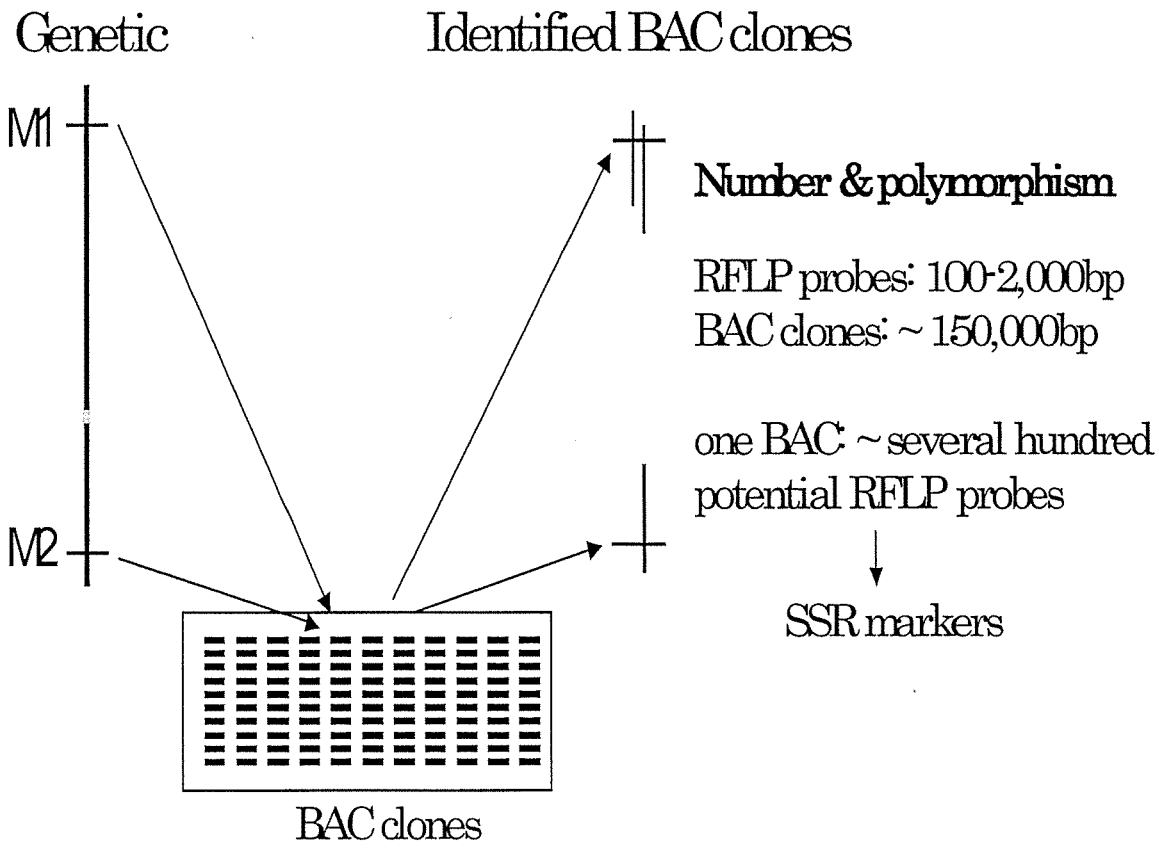


Figure 1.6: BAC technology in marker projects. BAC library is screened with molecular markers (M1 and M2). Positive BAC clones are identified and subcloned. Each BAC clone can generate several hundred RFLP probes and other types of markers.

cost. However, PCR-based markers, having a high multiplex ratio, make DNA marker technology more efficient and cost effective by striking a balance between cost and informativeness. With the development of PCR-based markers, saturated linkage maps have been constructed for all major species and it is possible to map and tag genes conditioning almost any trait. These maps have been extensively used in a wide spectrum of activities ranging from basic to applied research. This research has dramatically enhanced our understanding of the genetic bases of different characteristics and evolutionary relationships between species. The new knowledge gained offers an increased efficacy in species improvement and has also made it possible to design new strategies that allow the exploitation of genetic variation that was previously inaccessible.

The number of markers available for any given species is no longer a limiting factor in a marker project. But, even with the availability of the numerous marker systems, construction of a linkage map is still a time consuming and high-cost activity. For these reasons only one or a small number of linkage maps are available for any given species. Also, the number of genes segregating between any pair of genotypes is limited, so many different genotypes are involved in different research projects and the number can reach several thousand or more in a breeding program. Thus, markers and their mapping information need to be transferred to new populations constantly.

Marker transfer between populations/genotypes is by no means an easy task even for those from low-copy-number systems such as RFLP or SSR. One of the main reasons for this is that a marker would have to be able to detect difference(s) between the parents before it could be mapped, thus becoming useful. The chance for a marker to detect any difference between a pair of genotypes is generally low, although it varies depending on species, the closeness of the genotypes involved, as well as the marker

used. This low level of transferability is one of the major reasons why it is difficult to exploit these molecular markers in a breeding program.

The application of large-insert DNA libraries, BAC technology, offers the possibility to overcome the problem of low transferability, thus revolutionising marker research. With the use of existing mapping information, relevant large-insert clones could be isolated from a library. Due to their size differences, each of these large-insert clones can be utilised to generate hundreds of markers, all of which can be used to target the same genome region. Thus our ability to isolate polymorphic markers for a targeted region would be dramatically increased.

1.9 Research Objectives of This Study

The aims of this project included constructing a mungbean BAC library, and testing the feasibilities of exploiting the BAC library in generating polymorphic markers for gaps in linkage maps, in generating user-friendly markers for breeding programs, and in developing a set of BAC clones covering the mungbean genome for efficient generation of framework maps. Mungbean was selected for the study because, compared with other crop species, mungbean has a very small (579 Mbp) and highly polymorphic genome, thus it is highly suitable to test this or other new approaches.

Chapter 2

Construction of a Mungbean BAC Library

2.1 Introduction

The value of using large insert libraries in genome analysis has long been recognised. They have been extensively exploited in physical mapping (Zhang and Wing 1997), map based gene cloning (Zhang and Wing 1997), gene structure and function analysis of complex genomes (Zhang *et al.* 1996a), and in isolating molecular markers for targeted genome regions (Cregan *et al.* 1999).

There are three types of large-insert DNA libraries. One of these is the cosmid system (Hohn and Collings 1988). Inserts cloned in cosmid vectors are stable for long-term maintenance, and can be readily purified for manipulation and application. However, the cloning capacity of cosmid vectors is less than 50 kbp (Hohn and Collings 1988; Tao and Zhang 1998), which is not well suited for plant species, which often have huge genomes (Arumuganathan and Earle 1991). The second one is YAC (Yeast Artificial Chromosome) system (Burke *et al.* 1987). The dominant feature of this system is its huge cloning capacity, which can be over 1000 kbp in size (Burke *et al.* 1987; Chumakov *et al.* 1995; Libert *et al.* 1993; Umehara *et al.* 1995). However, a YAC insert can be very difficult to purify from the yeast host genomic DNA. The sizes of the 17 yeast chromosomes range from 210 to 1900 kbp, covering the size range of most YAC inserts (Zhang and Wing 1997). Further, chimerism, a clone containing DNA fragments from different parts of a genome, is another problem associated with the YAC system (Anderson 1993; Zhang and Wing 1997). The third one is the BAC system which was not reported until 1992 (Shizuya *et al.* 1992). This

system has overcome many of the problems associated with cosmid and YAC systems. Compared to cosmids, BAC contains much large inserts. It has been demonstrated that BAC can hold inserts of more than 300 kbp in size (Shizuya *et al.* 1992; Woo *et al.*1994). Compared to YAC, BAC libraries are easier to construct, maintain, and manipulate. It also has a very low level of chimaerism (Woo *et al.* 1994; Zhang *et al.* 1996b).

Due to its numerous advantages compared to cosmid and YAC systems, BAC has already become the system of choice for large DNA cloning of any organism with a large genome. Although it has only a short history, BAC libraries have been constructed for many species including *Arabidopsis* (Choi *et al.* 1995), lettuce (Frijters *et al.* 1997), sorghum (Woo *et al.* 1994), rice (Nakamura *et al.* 1997; Wang *et al.* 1995; Zhang *et al.* 1996b), apple (Vinatzer *et al.* 1998), melon (Luo *et al.* 2001), and soybean (Danesh *et al.* 1998; Tomkins *et al.* 1999a). However, there is no BAC or any other type of large DNA insert library available for mungbean yet. To make use of its overwhelming advantages, the BAC system was selected for this research. This chapter reports the construction and characterisation of a mungbean BAC library.

2.2 Materials and Methods

2.2.1 Plant Materials

Two mungbean genotypes, Accession 41 (ACC41) and ATF-3640, were used as the source of high-molecular-weight (HMW) DNA. Plants were grown in a glasshouse for 4 to 6 weeks at Samford in South East Queensland. Harvested leaf tissues were wrapped in aluminium foil and frozen in liquid nitrogen immediately. They were then stored in a –80°C freezer until use.

2.2.2 Isolation of High-Molecular-Weight DNA

Intact nuclei were isolated from leaf tissues following the protocol of Zhang *et al.* (1995). Approximately 50 g of frozen leaf tissues were ground into powder in liquid nitrogen with a cold mortar and pestle. The powder was immediately transferred into an ice-cold 2000 ml flask containing 500 ml ice-cold 1 x homogenization buffer (HB) [(10 x stock): 0.1 M trizma base, 0.8 M KCl, 0.1 M EDTA, 10 mM spermidine, 10 mM spermine, pH 9.4 - 9.5, plus 0.5 M sucrose and 0.5% Triton X-100]. ß-mercaptoethanol (0.15%) was added before use. The contents were gently swirled with a magnetic stir bar for 10 to 15 minutes on ice and filtered into six ice-cold 250 ml centrifuge bottles through two layers of miracloth and then another layer of miracloth by squeezing with gloved hands. The homogenate was pelleted by centrifugation with a fixed-angle rotor at 1800 *g* at 4°C for 20 minutes. The supernatant fluid was discarded and approximately 1 ml of ice-cold wash buffer (1 x HB plus 0.5% Triton X-100, 0.15% ß-mercaptoethanol was added before use) was added to each bottle. The pellet was gently resuspended with the assistance of a small paintbrush soaked in ice-cold wash buffer. The nuclei suspensions from all bottles were combined into one bottle and filled with ice-cold wash buffer. The nuclei were pelleted by centrifugation at 1800 *g* at 4°C for 15 minutes. The pellet was washed three additional times by resuspension in wash buffer and centrifugation at 1800 *g* at 4°C for 15 minutes. After the final wash, the pelleted nuclei were resuspended in a small amount (approximately 1 ml) of 1 x HB without ß-mercaptoethanol.

2.2.3 Embedding the Nuclei in Agarose Plugs

1% low-melting-point (LMP) agarose (BMA, USA) in 1 x HB (without ß-mercaptoethanol and Triton X-100) was prepared. Agarose was cooled to 45°C and maintained at this temperature in a water bath before use. The

nuclei were prewarmed to 45°C in a water bath and mixed with an equal volume of the 1% LMP agarose using a cut-off pipette tip. The mixture was aliquoted into ice-cold plug molds on ice with a cut-off pipette tip. When the agarose was completely solidified, the plugs were transferred to 5 to 10 volumes of lysis buffer (0.5 M EDTA, pH 9.0 - 9.3, 1% sodium lauryl sarcosine, and 0.3 mg/ml proteinase K). Agarose plugs were incubated in the lysis buffer for 36 to 48 hours at 50°C with gentle shaking. The plugs were washed once in 0.5 M EDTA, pH 9.0 - 9.3 for one hour at 50°C, once in 0.05 M EDTA, pH 8.0 for one hour on ice, and stored in 0.05 M EDTA, pH 8.0, at 4°C.

2.2.4 Pre-Electrophoresis of Agarose Plugs Containing HMW DNA

Agarose plugs were dialysed against sterile 0.5 x TBE buffer at least for 3 hours before pre-electrophoresis and then run in a 1% agarose gel (BMA, USA) using pulsed-field gel electrophoresis (PFGE, Bio-Rad, USA), with 0.5 x TBE buffer at 4 V/cm, 11°C, 5-s switch time for 6 hours. After electrophoresis, the plugs were collected from the wells and dialysed against 1 x TE buffer at 4°C for at least three hours with a change of TE buffer each hour. The treated plugs containing HMW DNA were stored at 4°C in TE buffer.

2.2.5 Partial Digestion of HMW DNA in Plugs

After dialysis of the pre-electrophoresed agarose plugs in TE buffer, one plug was divided into 12 mini-plugs of equal size. Mini-plugs were then aliquoted into six 1.5 ml microcentrifuge tubes, each containing two mini-plugs. 100 µl of pre-reaction buffer (1 x restriction buffer (Gibco BRL, USA), 2 mM spermidine (Sigma, USA), 0.1 mg/ml BSA and 1 mM DTT) was added into each tube and tubes were incubated on ice for 80 minutes with a change of buffer every 40 minutes. The DNA was digested with

two separate enzymes. The pre-reaction buffer was then replaced with reaction buffers that contained different concentrations of restriction enzymes: *Bam*HI (Gibco BRL, USA) or *Hind*III (Gibco BRL, USA) ranging from 0.0 units to 1.5 units. The reaction mixtures were incubated on ice for 60 minutes to allow the enzyme to access the DNA in the agarose mini-plugs. The reaction mixtures were then transferred to a 37°C water bath and incubated for 10 minutes. The reactions were immediately stopped by adding one-tenth volume of 0.5 M EDTA. The agarose mini-plugs were loaded into a 1% agarose gel (PROGEN, Australia) and sealed in position with 1% LMP gel. Low-range PFG marker (New England BioLabs, USA) was applied to flanking wells. PFGE was performed at 6 V/cm, 11°C, 90-s switch time in 0.5 x TBE buffer for 18 hours. The gel was stained in ethidium bromide for 20 minutes and then rinsed with water. The gel was photographed and the enzyme concentration under which most of the partially restricted fragments fell into the range from 100 to 400 kbp was used for large-scale partial digestion, which was carried out using ten plugs.

2.2.6 Size-Selection for Partially Digested DNA Fragments

The agarose mini-plugs containing partially digested HMW DNA were applied to the centre of a 1% agarose gel (BMA, USA) and low-range PFG marker was applied to flanking wells. The size fractionation was performed in a PFGE apparatus under conditions described above. The flanking marker lanes were removed from the gel and stained with ethidium bromide to indicate the location of the size ranges. Gel slices were then cut from the genomic DNA lanes by horizontal cuts at 0.5 cm intervals to obtain gel slices in the range of 100 to 400 kbp. The size-fractionated DNA in the agarose slices was stored in 0.5 M EDTA (pH 8.0) at 4°C until use.

2.2.7 Second Size Selection and Recovery of HMW DNA

The agarose slice containing size-fractionated DNA was put in dialysis tubing, 3/4 inch in diameter, (Gibco BRL, USA) with an appropriate volume of 0.5 x TBE buffer, which just covered the gel. The DNA was electroeluted at 6 V/cm, 11°C, 30-s switch time in 0.5 x TBE for 4 hours followed by reversing the direction of dialysis tubing at 180° for 90 seconds to release DNA from the wall of the bag. The DNA was transferred to a 1.5 ml microcentrifuge tube with a wide-bore pipette tip. Loading buffer was added to eluted DNA, which was subjected to the second round of size selection in 1% agarose gel with low-range PFG marker in flanking wells at 4 V/cm, 11°C, 5-s switch time in 0.5 x TBE buffer for 8 hours. The flanking marker lanes were removed from the gel and stained with ethidium bromide to indicate the location of the DNA fragments in the gel. The gel slice containing second size selected DNA was cut from the gel. The DNA was finally recovered by electroelution as above, but in 1/4 inch dialysis tubing, and dialysed against TE (pH 8.0) for 3 hours at 4°C with a change of TE buffer each hour. The DNA was removed gently using a wide-bore pipette tip and transferred to a clean microtube. The DNA concentration was examined using electrophoresis with lambda DNA as a standard.

2.2.8 Preparation of BAC Vector

Low-copy pBeloBAC11was used as the cloning vectors for the construction of the BAC libraries and was isolated using the Plasmid Maxi Kit (Qiagen, Germany). Stock cells of F-factor-based plasmid BAC vector, pBeloBAC11 (7.4 kbp), were streaked on an LB plate containing 12.5 µg/ml chloramphenicol and grown overnight at 37°C. A single colony was picked from a freshly streaked plate and inoculated a starter culture of 5 ml LB medium plus 12.5 µg/ml chloramphenicol in a 50 ml Falcon tube and

grown at 37°C for 8 hours with vigorous shaking (300 rpm). LB medium (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% sodium chloride, pH 7.5) containing 12.5 µg/ml chloramphenicol was prepared in two 2-litre flasks containing 1 litre of medium in each flask inoculated with 1 ml of the starter culture. The culture was grown at 37°C for 12 to 16 hours with vigorous shaking (300 rpm). Cells were harvested in ice-cold 250 ml centrifuge bottles by centrifugation at 6000 *g* for 15 minutes at 4°C and supernatant fluid was discarded. The bacterial pellet was completely resuspended in a small amount of BufferP1 and additional BufferP1 was added to the appropriate volume (100 ml of BufferP1 for 2 litres of culture). The appropriate volume (100 ml) of BufferP2 was added and mixed gently by inverting bottles 4 to 6 times and incubating at room temperature for 5 minutes. Chilled BufferP3 (100 ml) was added and mixed immediately, but gently, by inverting bottles 4 to 6 times and incubating on ice for 30 minutes with gentle mixing every 10 minutes. The sample was centrifuged at 20,000 *g* for 30 minutes at 4°C and supernatant containing plasmid DNA was transferred into clean bottles. The supernatant was re-centrifuged at 20,000 *g* for 15 minutes at 4°C and then transferred into new bottles. DNA in supernatant was precipitated by mixing with 0.7 volumes of room-temperature isopropanol. The mixture was then centrifuged at 15,000 *g* for 30 minutes at 4°C. The supernatant was carefully decanted. The DNA pellet was redissolved in 1 ml TE, pH 7.0, and left for 30 minutes on the bench. Buffer QBT was added to obtain a final volume of 12 ml. A QIAGEN-tip 500 was equilibrated by applying 10 ml of Buffer QBT and allowing the column to empty by gravity flow. The DNA solution was applied to the QIAGEN-tip 500 and allowed to enter the resin by gravity flow. QIAGEN-tip was washed twice with 30 ml of Buffer QC. DNA was eluted with 15 ml of Buffer QF and the DNA was precipitated by adding 0.7 volumes of room-temperature isopropanol. The sample was mixed and immediately centrifuged at 15,000 *g* for 30 minutes. The supernatant was carefully decanted. The DNA pellet was

washed with 2 to 10 ml room-temperature 70% ethanol and centrifuged at 15,000 *g* for 10 minutes and the supernatant was carefully decanted. The pellet was air-dried and redissolved in TE, pH 8.0 buffer. The DNA concentration was determined comparison of band intensity to lambda DNA molecular weight markers on agarose.

The pBeloBAC11 DNA was digested with 3 units of either *Hind*III or *Bam*HI per microgram DNA at 37°C. After 2 hours of incubation, an additional 2 units of enzymes per microgram DNA were added and the incubation continued for additional 2 hours. To test the efficacy of the reaction, some of the reaction mix was run on an agarose gel. The digest was then subjected to phenol:chloroform:isoamylalcohol (25:24:1) and chloroform purification. Purified DNA solution was then precipitated with one tenth volume of 3 M sodium acetate (pH 5.2) and 2 volumes of 100% cold ethanol. After washing the pellet twice with 70% ethanol, followed by centrifuging for 30 minutes at 4°C twice, the pellet was dried on a bench and dissolved in water. The ends of the linearised pBeloBAC11 DNA were de-phosphorylated with SAP (shrimp alkaline phosphatase, Roche, Germany) according to the manufacturer's specification. After de-phosphorylation, the phosphatase was inactivated at 65°C for 15 minutes. The vector was stored in 10 µl aliquots in a –20°C freezer.

2.2.9 Ligation and Transformation

Ligation and transformation were performed using the protocol described by Osoegawa *et al.* (1998). The ligation was carried out in 50 µl reactions, in which the molar ratio of electroeluted DNA to either *Bam*HI-digested or *Hind*III-digested and de-phosphorylated pBeloBAC11 vector at approximately 1:8, 1:10 and 1:12 molar ratios of insert : vector. Ligation was carried out at 16°C overnight. The reaction was stopped by incubation with 2 µl EDTA (pH 8.0, 0.5 M) and 1 µl Proteinase K (10 mg/ml) at 37°C

for 1 hour, followed by the addition of 1 µl PMSF (phenylmethylsulfonyl fluoride, from Sigma, 100mM in isopropanol) to inactivate residual Proteinase K. The ligation mixture was spotted on the middle of a microdialysis filter (0.025 µm pore size; Millipore) floating on sterile deionised water in a disposable Petri dish, and dialysis was continued for at least 2 hours on ice. The dialyzed solution was recovered carefully from the membrane with a cut-off pipette and the membrane was transferred onto 0.5 x TE containing 30% PEG8000. The solution was placed on the membrane again, and dialysis was continued for approximately 5 hours until equilibrated.

Escherichia coli strain DH10B was used as the host. The ligated DNA was transformed into ElectroMAX DH 10B cells by electroporation using a Cell Porator and Voltage Booster system (Gibco BRL, USA) as specified by the cell provider (Gibco BRL, USA). The ligation product (2 µl) was added to 20 µl of the cells for a single electroporation. The Cell Porator settings were 350 V, 330 µF capacitance, low ohms impedance and fast charge rate, and the Voltage Booster setting was 4 K ohms resistance. After electroporation, the cells were collected from the cuvette and resuspended in 1 ml SOC medium (2% bacto-tryptone, 0.5% bacto-yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose pH 7.0) and incubated at 37°C for 1 hour with shaking at 200 rpm to allow the cells to express an antibiotic (chloramphenicol) resistance gene. The cells were then plated on the LB plate (1.5% agar containing 12.5 µg/ml chloramphenicol, isopropyithio-ß-D-galactoside (IPTG, 200mg/ml) and 5-bromo-4-chloro-3-indolyl-ß-galactoside (X-gal, 20mg/ml) and grown at 37°C for 24 to 36 hours. The recombinant clones (BACs) were clearly identified by a blue (nonrecombinant) or white (recombinant) phenotype. White colonies were manually picked and stored in 384-well microtiter plates, each well containing 70 µl of LB with 12.5 µg/ml chloramphenicol

and 15% glycerol. Bacteria were grown for 24 hours at 37°C. Each of the 384-well microtiter plates containing grown bacteria were replicated using a 384-pin hand replicator and placed at –80°C for long-term storage. Bacteria in replicates were grown for 24 hours at 37°C as per originals and placed at –80°C freezer for long-term storage.

2.2.10 Analysis of BAC DNA

Randomly selected BAC clones were inoculated in 5 ml of LB broth containing 12.5 µg/ml chloramphenicol and grown at 37°C with shaking at 250 rpm for 16 to 20 hours. 100 µl of each overnight culture was transferred into 1.5 ml microcentrifuge tubes and glycerol was added to final concentration of 15%. Tubes were stored at –80°C as permanent stocks. The circular BAC DNA was isolated from the rest of the overnight cultures by the alkaline lysis method of Sambrook *et al.* (1989) used for the preparation of plasmid DNA, with the following modifications. Bacterial cultures were centrifuged at 12,000 *g* for 10 minutes and bacterial pellets were resuspended in ice-cold Solution I (50 mM glucose, 25 mM Tris HCl, pH 8.0, 10 mM EDTA, pH 8.0) by vortexing, and then incubated on ice for 5 minutes. Freshly prepared Solution II (0.2 N NaOH, 1% SDS) was added into the bacterial cultures and tubes were incubated on ice for 5 minutes. Solution III (3 M potassium acetate, 5 M glacial acetic acid) was then added and mixed gently, cultures were frozen at –80°C for 15 minutes and thawed at room temperature. Cultures were centrifuged at 2800 *g* for 15 minutes in a desktop centrifuge and 0.75 ml of the supernatant was transferred into 1.5 ml microcentrifuge tubes containing 0.45 ml ice-cold isopropanol. Contents were mixed and centrifuged at 12,000 *g* in a microcentrifuge for 5 minutes to pellet DNA. The supernatant was removed and pellets were washed with 200 µl of ice-cold 70% ethanol by centrifugation at 12,000 *g* for 2 minutes. Ethanol was removed with a pipette and DNA pellets were dried on a bench. DNA was

dissolved in 40 µl TE (pH 8.0). Approximately 3 µg of BAC DNA was digested with 3 units of *Not*I restriction enzyme in a 40 µl volume (1 x *Not*I buffer, 2 µM spermidine) for at least 3 hours at 37°C to release the mungbean DNA insert from the cloning vector. The reaction was stopped by adding one-tenth volume of 6 x loading dye. The digested DNA was analysed by PFGE in 1% agarose, with a low-range PFG marker in flanking wells in 0.5 x TBE buffer at 11°C and 6 V/cm for 16 hours with 5-s to 15-s pulse time at a 120° angle. The gel was stained for 30 minutes with ethidium bromide and destained for 30 minutes in water. The gel was photographed to determine the average insert size.

Once an average insert size was calculated, the probability of recovering any specific sequence of interest was calculated by using the following equation (Clarke and Carbon 1976):

$P = 1 - \{1-(L/G)\}^{N}$ where,

N = number of clones in library

P = probability to get the target gene

L = length of average clone insert in basepairs

G =haploid genome length in basepairs (Genome sizes are from Arumuganathan and Earle 1991)

2.2.11 Colony Filters

The mungbean library stored in 384-well microtiter plates was replicated using a 384-pin hand replicator onto Hybond N + membrane, 7.5 cm x 11.5 cm (Amersham). The inoculated filters were placed on LB plates containing 12.5 µg/ml chloramphenicol and incubated at 37°C until the colonies reached a size of 2 to 3 mm in diameter (14 to 16 hours). The colony filters were prepared following the methods as described by Zhang *et al.* (1996b). The filter was placed colony side up onto three layers of Whatman 3 MM paper saturated with the following solutions, incubated

at room temperature, and transferred in order: (1) 10% SDS, 4 minutes; (2) Denaturing solution (0.5 M Tris. HCl, pH 8.0, 1.5 M NaCl, 1 mM EDTA, pH 8.0), 5 minutes; (3) Neutralising solution (0.5 N NaOH, 1.5 M NaCl), 2 x 5 minutes; (4) 2 x SSC, 0.1% SDS, 5 minutes; (5) 2 x SSC, 5 minutes; (6) 0.4 N NaOH, 20 minutes. The filter was then washed twice in 5 x SSC, 0.1% SDS, 20 minutes each time. The filter was then rinsed twice in 2 x SSC, 10 minutes. Membranes were blotted on papers to remove all excess liquid, kept in plastic bags, and stored at –20°C.

2.2.12 RFLP Clones

All RFLP probes used in this study were *Pst*I clones constructed using total genomic DNA from leaves of mungbean cultivars ' Berken' and prefixed by VrCS, where 'Vr' represents the species name *Vigna radiata,* and 'CS', CSIRO, the institution where the research was carried out (Humphry *et al.* 2002). These clones were used in construction of a mungbean linkage map and the copy numbers were previously analysed between 'Berken' and 'Acc41' by Humphry *et al.* (2002).

Eight single-copy number RFLP clones were grown in 5 ml of LB/ampicillin (0.1 mg/ml) overnight. Plasmid DNA was isolated by the alkaline lysis method and the DNA solution (500 to 1000 ng) was digested with a restriction endonuclease *Pst*I in a total volume of 30 µl, consisting of 10 units of *Pst*I (Promega), 1x *Pst*I buffer, 2 µM spermidine, 0.1 µg/µl bovine serum albumin (BSA) and 0.4 µg/µl RNase (Qiagen) at 37°C overnight. The reaction was stopped by adding loading buffer.

Following digestion, the inserts were separated from the vectors by electrophoresis through a 1% agarose gel in 0.5x TBE buffer at 100 V for 3-5 hours. DNA fragments containing the inserts were excised from the gel

and transferred into 1.5 ml microcentrifuge tubes. Inserts were purified using the Qiaex II Agarose Gel Extraction Kit according to the manufacturer's instructions (Qiagen). Purified DNA was quantified on a gel against molecular weight markers.

2.2.13 Mungbean BAC Library Screening Procedure

Colony filters were hybridised using standard techniques as described by Sharp *et al.* (1988). Prehybridisation of the nylon membrane prior to probe labeling was carried out for 5 to 6 hours at 65°C in a small box containing a set of the entire colony filters. 200 ml of a buffer solution containing: 10 x Denhardt's III [2% gelatin, 2% ficol-400 (Sigma F-4375), 2% PVP-360 (Sigma PVP-360), 10% SDS, 5% Sodium pyrophosphate], one- fifth volume 5 x HSB [17.53% NaCl, 3.03% PIPES (Sigma P-6757), 0.745% EDTA] and 5 mg/ml salmon sperm DNA was used. The volume of the buffer was reduced to just cover membranes prior to addition of the probes.

A commercially available labelling kit, Amersham Megaprime, (Amersham, Australia) was used for probe labeling. Screening with eight single-copy mungbean RFLP probes was performed to evaluate the feasibility of detecting positive BAC clones from the library and to assist the estimation of the genome coverage of the collected BAC clones. Approximately 300 ng (6 µl) of probe was mixed with 4 µl of primer buffer and 9 µl of water and then placed into a 1.5 ml microcentrifuge tube. The tube was boiled for 5 minutes to denature DNA. After boiling, the tube was incubated on ice for 2 minutes and then 6 µl of labelling buffer and 2 µl of Klenow fragment were added. The tube was centrifuged briefly and 2 µl of α-^{32}PdCTP (Perkin-Elmer, USA) was added. The tube was incubated at 37°C for 2 hours. To denature the labeled probe DNA, 5 µl of 4 N NaOH was added into the tube and the tube was then incubated for 3 minutes at room temperature. Labelled probe was then used for hybridisation which

was carried out at 65°C for 18 to 24 hours.

Filters were washed three times at 65°C prior to exposure on Kodak X-Omat film (Kodak Eastman). The first and second washes were done in solutions containing 2 x SSC and 1% SDS, and the last wash was done with a solution containing 0.2 x SSC and 1% SDS.

2.3 Results

2.3.1 Partial Digestion

Results of partial digestions with both *Bam*HI and *Hind*III are shown in Figure 2.1. As shown in the figure, over-digestion was apparent with the use of 1.5 units of *Hind*III. For this restriction enzyme, the use of between 0.4 and 0.8 units produced DNA fragments of the desired size (150-400 kbp). For *Bam*HI, over-digestions were apparent when more than 1.5 units of the enzyme was used. Thus, large-scale digestions of HMW DNA were performed using 0.4 to 0.8 units of *Hind*III and 0.1 to 0.2 units of *Bam*HI.

2.3.2 Construction of ACC41 and ATF-3640 BAC Libraries

Two mungbean BAC libraries were constructed, one from ACC41 and the other from ATF-3640. The ACC41 BAC library contained 6912 clones, and was constructed using *Hind*III partially digested DNA fragments that were size-selected twice on pulsed-field gels. The ATF-3640 BAC library contained 11904 clones, and was constructed from the *Bam*HI partially digested fragments that were size-selected only once on a

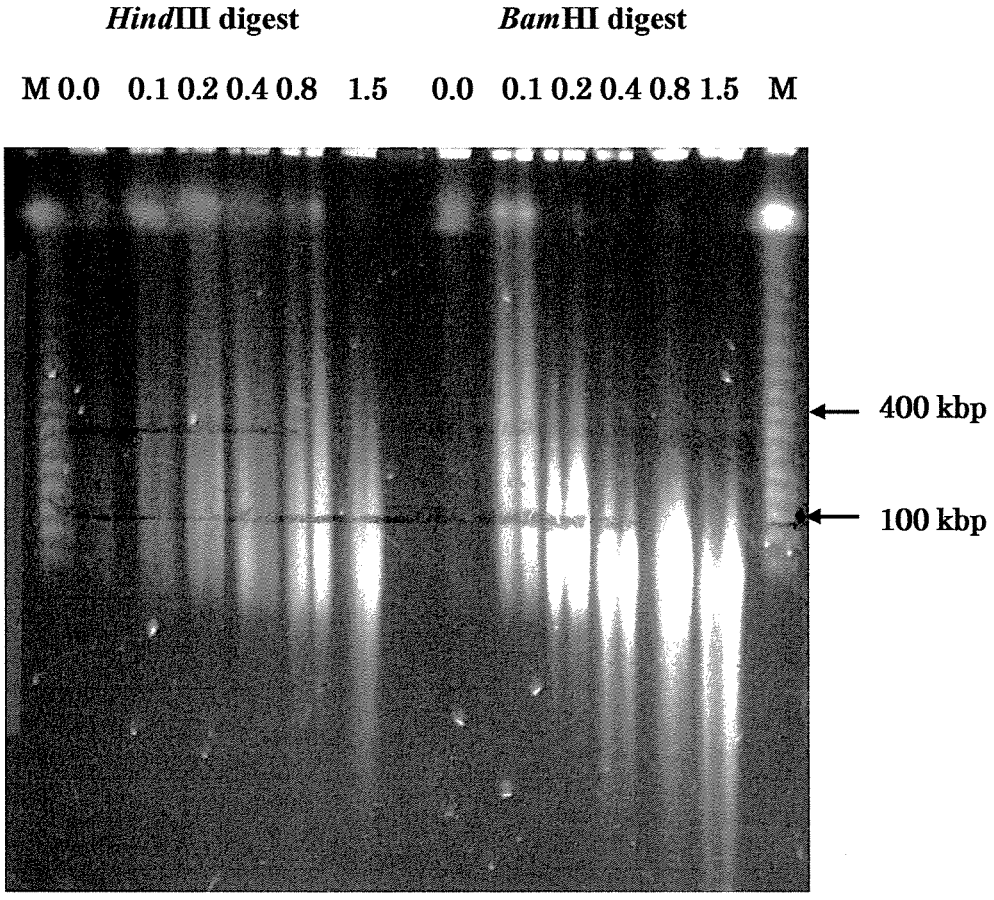


Figure 2.1: Results of partial digestion of HMW mungbean DNA. (M): low-range PFG marker, Lanes 2-7: mungbean DNA partially digested with *Hind*III (0.0) 0 U (0.1) 0.1 U (0.2) 0.2 U (0.4) 0.4 U (0.8) 0.8 (1.5) 1.5 U. Lanes 9-12: mungbean DNA partially digested with *Bam* HI (0.0) 0 U (0.1) 0.1 U (0.2) 0.2 U (0.4) 0.4 U (0.8) 0.8 (1.5) 1.5 U. The genomic DNA was subjected to PFGE on a 1% agarose in 0.5 x TBE using a switch time of 90-s at 6 V/cm and 11°C for 18 hours.

pulsed-field gel. These two separate ligation experiments yielded a combined mungbean BAC library consisting of 18,816 clones. These clones were individually stored in 49 384-well microtitre plates in a –80°C freezer.

2.3.3 Characterisation of the BAC Libraries

2.3.3.1 Insert Size Distribution of the Clones from the ACC41 and ATF-3640 BAC Libraries

To determine the distribution of BAC insert sizes and the average insert sizes of the two BAC libraries, DNA was isolated from 61 random BAC clones from the ACC41 library and 82 random clones from the ATF-3640 library. The 61 BAC clones from the ACC41 library contained inserts ranging from 50 to 180 kbp, with an average length of 112 kbp (Figure 2.2). The distribution of the insert sizes is shown in Figure 2.3. The 82 random clones from the ATF-3640 library contained inserts ranging from 15 to 185 kbp, with an average length of 102 kbp (Figure 2.4). The distribution of inserts from this library is shown in Figure 2.5.

2.3.3.2 Screening the ACC41 and ATF-3640 BAC Libraries with Mungbean RFLP Clones

To test the feasibility of isolating BAC clones from the two libraries, eight single-copy-number RFLP clones were screened against the combined 18,816 mungbean BAC clones. These eight probes identified between 2 to 6 positive BAC clones each (Table 2.1). An example of an autoradiogram image of the BAC colony filters of the mungbean BAC library hybridised with a mungbean single-copy RFLP probe is shown in Figure 2.6.

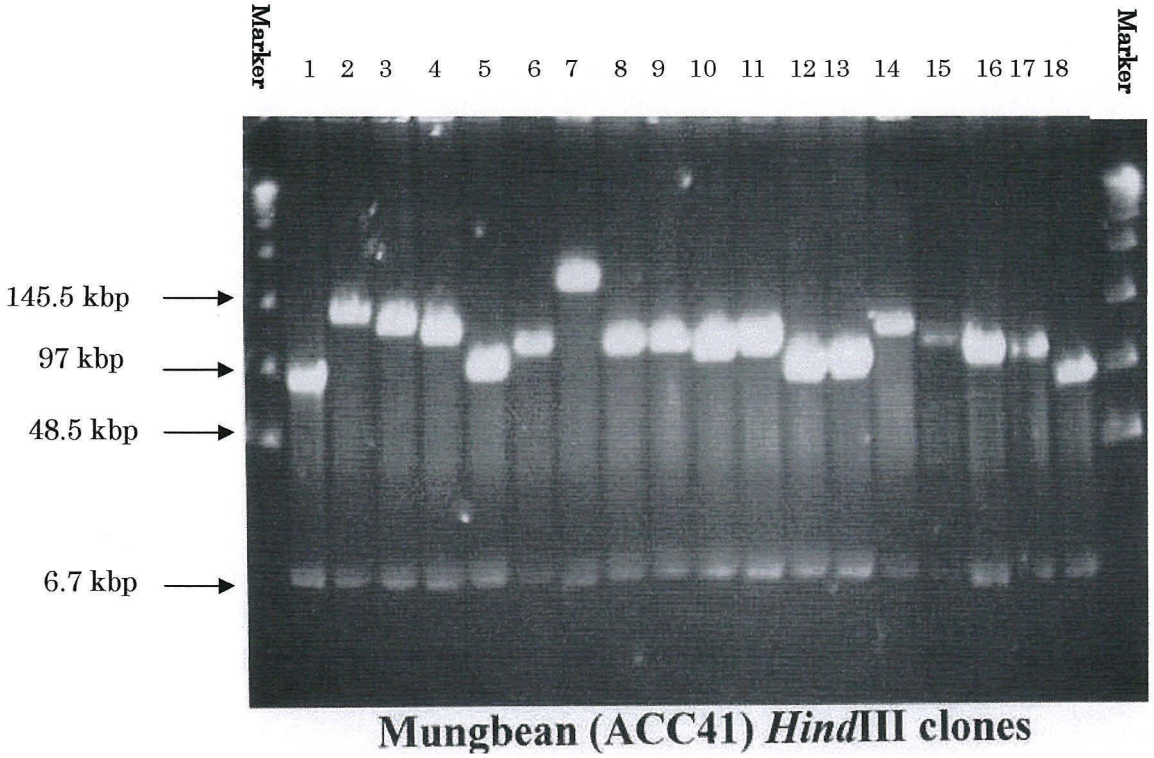


Figure 2.2: Insert sizes of 18 random *Hind*III clones from the ACC41 library. Lanes 2-18 are alkaline lysis minipreps of recombinant BAC clones digested with *Not*I. The 6.7 kbp band in each lane is pBeloBAC11. PFGE was performed with a CHEF DRIII (BioRad, USA) under the following conditions: 1% agarose gel in 0.5 x TBE, T = 11°C, V = 6 V/cm, pulse time = 5-s to 15-s at a 120° angle, and run = 18 h.

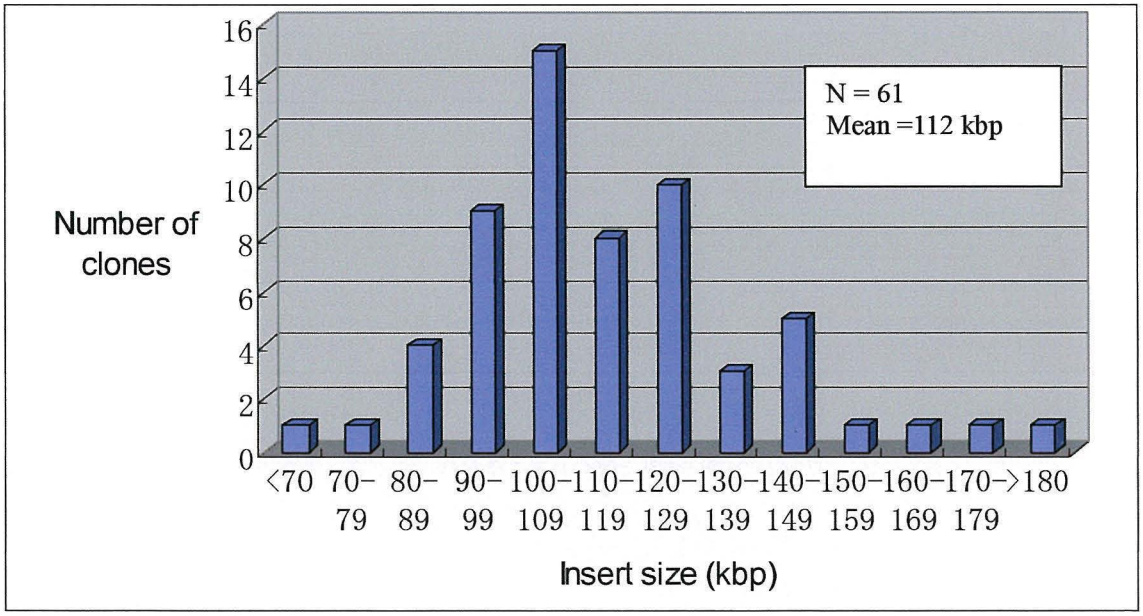


Figure 2.3: Insert size distribution of the 61 BAC clones randomly taken from the ACC41 BAC library. To determine the size distribution of BAC in the ACC 41 BAC library, inserts of the 61 BAC clones were plotted against the frequency of each group of clones represented in the library.

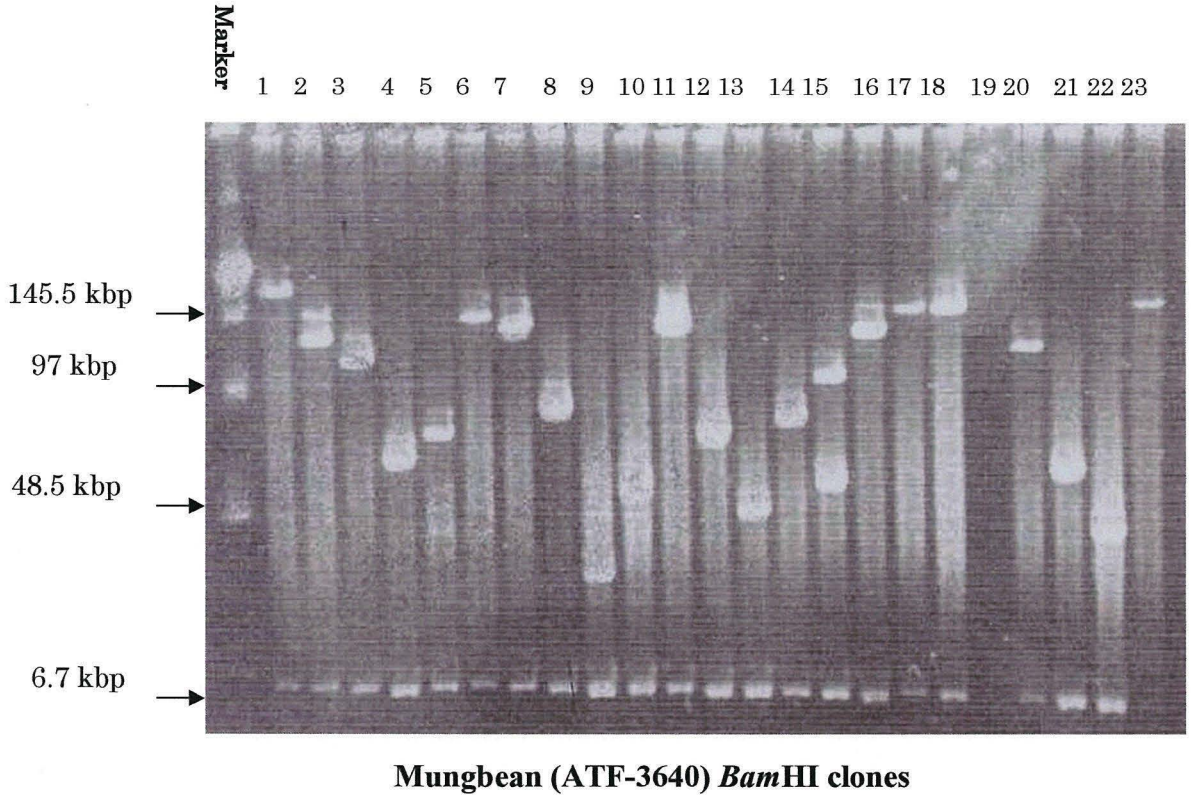


Figure 2.4: Analysis of 23 random *Bam*HI clones from the ATF-3640 library. Lane 1 is molecular-weight lambda concatemer. Lanes 2-23 are alkaline lysis minipreps of recombinant BAC clones digested with *Not*I. The 6.7 kbp band in each lane is pBeloBAC11.

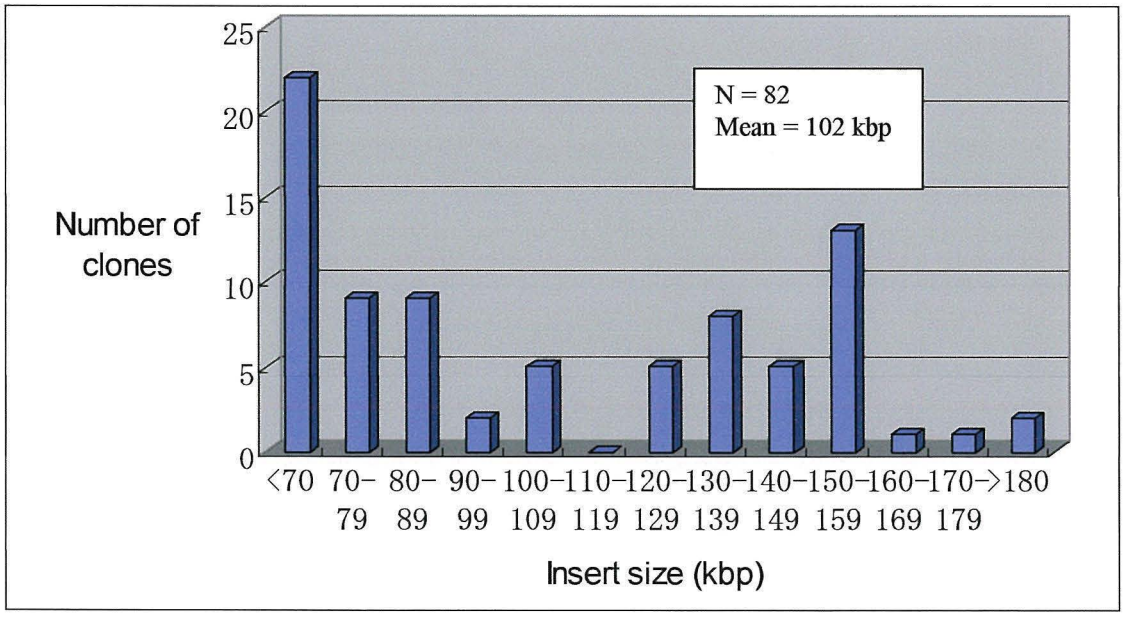


Figure 2.5: Insert size distribution of BAC clones randomly taken from the ATF-3640 BAC library.

Table 2.1: Positive BAC clones identified by eight single-copy RFLP clones. LG = linkage groups in mungbean linkage map (Humphry *et al.* 2002)

Probe	LG	Positive BAC clones
VrCS13	D	8L2, 8O14, 22C3, 32N19
VrCS19	H	2H2, 24O16, 35E16, 38J22, 40G21, 43I7
VrCS66	E	47H13, 49E5
VrCS126	A	39G18, 39N21
VrCS176	G	38H14, 38J13, 44K6, 47G15
VrCS198	E	26A4, 31P13, 44A6
VrCS296	K	23O5, 41O2, 27L11, 50K15
VrCS375	L	24B4, 29N9, 33D20, 34J8, 37G14, 38L19

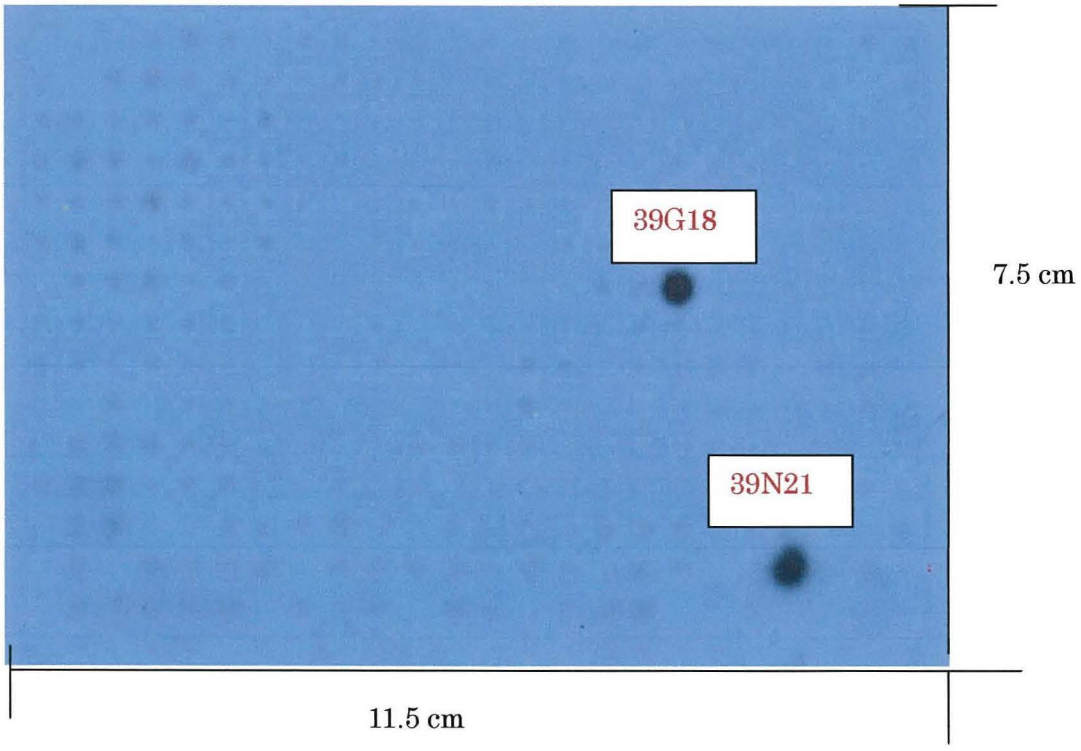


Figure 2.6: An autoradiogram image showing two (39G18 and 39N21, as marked) positive BAC clones identified by probe VrCS126.

2.4 Discussion

2.4.1 Genome Coverage of the Mungbean BAC Libraries

Two mungbean BAC libraries were constructed in this study. The first one consisted of 6912 *Hind*III clones from the genotype ACC41. These clones had an average insert size of 112 kbp. The second library consisted of 11904 *Bam*HI clones from the genotype ATF-3640. These clones had an estimated average insert size of 102 kbp. Thus, the combined clones have a total of approximately 2018 Mbp of nuclear DNA. Considering that the genome size of mungbean is 579 Mbp (Arumuganathan and Earle 1991), these BAC clones represent 3.56 genome equivalents. Based on the calculation of Choi and Wing (2000), the probability of getting a specific sequence at least once from the combined BAC clones is approximately 96%. These estimations were supported by the results from screening mungbean BAC clones with eight single-copy RFLP probes. They detected two to six positive BAC clones each (Table 2.1). Further, results from previous studies indicated that the use of a single restriction enzyme may end up with preferencial cloning (Choi and Wing 2000), thus the use of two different restriction enzymes in here may also enhance the probability of genome coverage.

2.4.2 Effect of Second Round of Size Fractionation in Library Construction

Previous studies have indicated that a second round of size fractionation often results in BAC clones with more even insert sizes, although this is often accompanied by decreased transformation efficiency which is likely caused by the increased insert sizes (Zhang *et al.* 1996b). This study supports these results. A second round of size-selection for the construction of the ACC41 mungbean library not only resulted in an

increase in average insert size (by 10 kbp), but also produced clones with much more even insert sizes (50 – 180 kbp compared to 15-185 kbp from the single size fractionation). Further, differences in the transformation efficiencies between the two libraries were negligible. Thus, it seems that it is highly desirable to have a second round of size fractionation in BAC construction.

2.4.3 Usefulness of the Mungbean BAC Libraries

BAC libraries have been extensively exploited in many research areas including generating high-resolution physical maps and identifying candidate genes during physical cloning (Arondel *et al.* 1992; Tanksley *et al.* 1995), to study the relationship between physical and genetic distance (Civardi *et al.* 1994), and to assess genome architecture (San Miguel *et al.* 1996). Thus the first mungbean libraries constructed in this study could find applications in all these different areas. Considering the relatively small genome size of the mungbean genome (579 Mbp, Arumuganathan and Earle 1991), the BAC libraries would be especially useful in research associated with physical mapping and map-based gene cloning.

The main interest in these BAC clones is to develop polymorphic markers for targeted genome regions. Considering the size differences between molecular markers (up to 2 kbp in size) and BAC clones (over 100 kbp in average), a large number of molecular markers can potentially be isolated from each BAC clone. Previous studies have found that the frequency of di- or tri- or tetra-nucleotide microsatellites in plant genomes is one in every 21.2 kbp in dicotyledonous plants (Wang *et al.* 1994). Thus, on average, several SSRs could be generated from each of the BAC clones. Therefore, the BAC libraries constructed should be more than adequate for this research project.

Chapter 3

Efficient Generation of Polymorphic Markers for Specific Chromosomal Regions and Framework Maps for New Mapping Populations by Exploiting BAC (bacterial artificial chromosome) Clones

3.1 Introduction

Linkage maps are powerful tools in modern genetics and they have been constructed for all major organisms. These maps have been intensively exploited in a wide spectrum of activities ranging from basic to applied research. These activities have dramatically enhanced our understanding of the genetic bases of different characteristics (Buerstmayr *et al.* 2002; Lippman and Tanksley 2001) and of evolutionary relationships between species (Gale and Devos 1998b; Paterson *et al.* 1995). Molecular markers for a wide range of characteristics have been developed and their application has the potential to increase the predictability of breeding programs.

Over the last decade or so, numerous molecular marker systems have been developed. However, construction of a linkage map with good genome coverage is still a time consuming and costly activity. In spite of extensive international effort, we have only a limited number of linkage maps for any given species. In contrast, the number of genes segregating between any pair of genotypes is limited, so many different genotypes are involved in different research projects and the number can reach several thousands or more in a breeding program. To locate genes conditioning a trait or to develop markers for a gene, linkage maps for new populations that segregate for the traits of interest are often required. However,

irrespective of species and marker systems used, linkage maps constructed with existing techniques often contain gaps (Chalmers *et al.* 2001; Messmer *et al.* 1999; Ulloa *et al.* 2002). These gaps do not allow a whole genome to be scanned for identifying genes underlying a trait.

Michelmore *et al.* (1991) developed a method (Figure 3.1) to generate polymorphic markers for specific chromosomal regions and the method has been successfully used (Campbell *et al.* 2001; Kiehne and Neale 1998). This method is based on comparing two pooled DNA samples by the use of random markers. To produce the two DNA pools, markers flanking the target region are required. However, such markers are not always available. These include situations where no mapped markers are available for a large segment of a chromosome or even a whole chromosome (Chalmers *et al.* 2001). Further, this method uses random markers and it may require the testing of a large number of markers before a polymorphic marker for a specific chromosome region can be found (Ardiel *et al.* 2002; Roy *et al.* 1999).

Another possible method to develop markers for specific chromosomal regions is by transferring markers with known locations from existing maps to new populations. This method, however, can be very difficult because many markers from a linkage map would not be able to detect polymorphism between the parents of a new population. The proportion of markers that could detect polymorphism depends on species as well as the distance between parents of a particular population. It is interesting to note that, to allow a maximum number of markers to be mapped, populations of wide crosses have been purposely selected in linkage map construction for many species (Messmer *et al.* 1999; Röder *et al.* 1998). Thus the proportion of markers that could be utilized in a population derived from an "average" cross could be much lower. The low level of marker transferability is further complicated by the fact that many

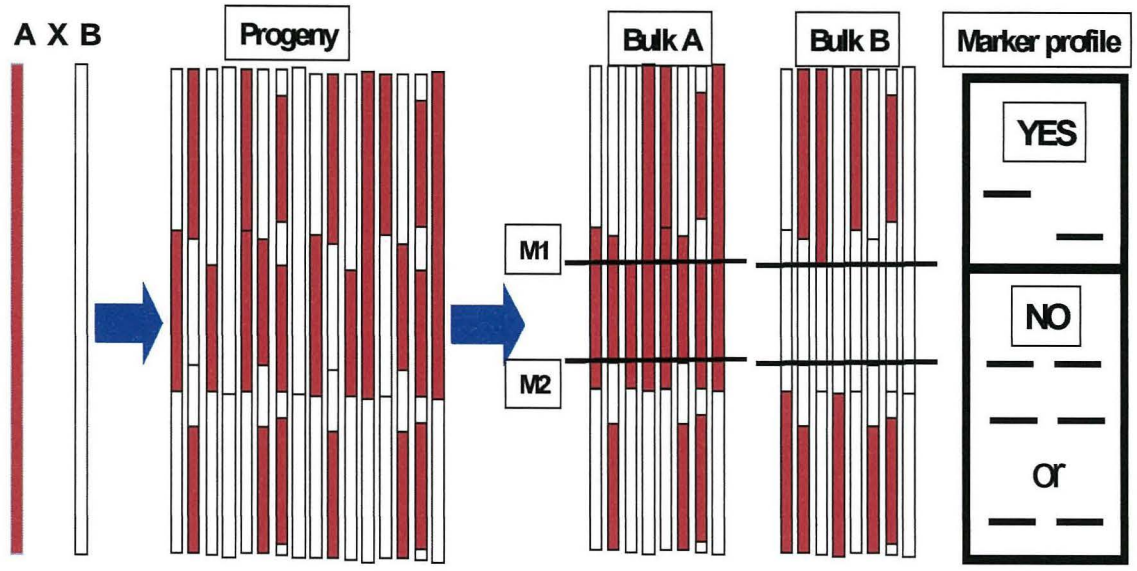


Figure 3.1: BSA (Bulked Segregation Analysis, Michelmore *et al.* 1991). This method requires a segregating population. Based on the genotype of the markers flanking a target region, individuals, homozygous for parental alleles across the target interval, are selected from the segregating population. These form the two bulks, A and B, then markers are checked against the two bulks. If a marker shows two different profiles, it indicates the marker is likely to be located between the flanking markers. In other words, the marker is likely to be located in the target region. If the two bulks produce the same profile for a marker, it indicates that the marker is not located in the target region. The exact position of the marker can then be determined by segregation analysis.

linkage maps have marker clusters in some genome regions while having few markers in others (Messmer *et al.* 1999). The available markers for some genome regions can be limited, thus the chance to get polymorphic markers for a new population for these chromosome regions could be very low.

It is of interest to note that often a large number of markers are needed for genome mapping and gene tagging. However, the number of markers alone does not guarantee a map with good coverage. The distribution of the markers across the genome, or genome coverage, is more relevant. Similarly, a large number of markers will not necessarily allow the isolation of markers closely linked to a particular gene of interest, but the locations of these markers are more important. In this chapter, a new approach that offers the potential for more efficient isolation of polymorphic markers for specific chromosomal regions, and for more efficient generation of framework maps, is demonstrated (Figure 3.2). The approach can also be easily adapted to generate markers closely linked to a gene when loosely linked markers are available. The feasibility of this approach was demonstrated in mungbean (*Vigna radiata*) using a chromosomal region harbouring a major locus conditioning powdery mildew (PM) resistance as the target.

To study the genetics of resistance to powdery mildew in mungbean, 322 RFLP clones, including all those that detected the 260 loci of an existing map (Humphry, unpublished), were screened against the PM parents. Even with the use of 10 restriction enzymes, polymorphic markers could only be found to cover less than 50% of the genome, leaving many gaps including three whole linkage groups that did not have any markers. However, mapping the polymorphic markers against the PM population did locate a major gene conditioning powdery mildew resistance between XvrCS296 and XvrCS73 on linkage group K (Humphry *et al.* 2002). Six

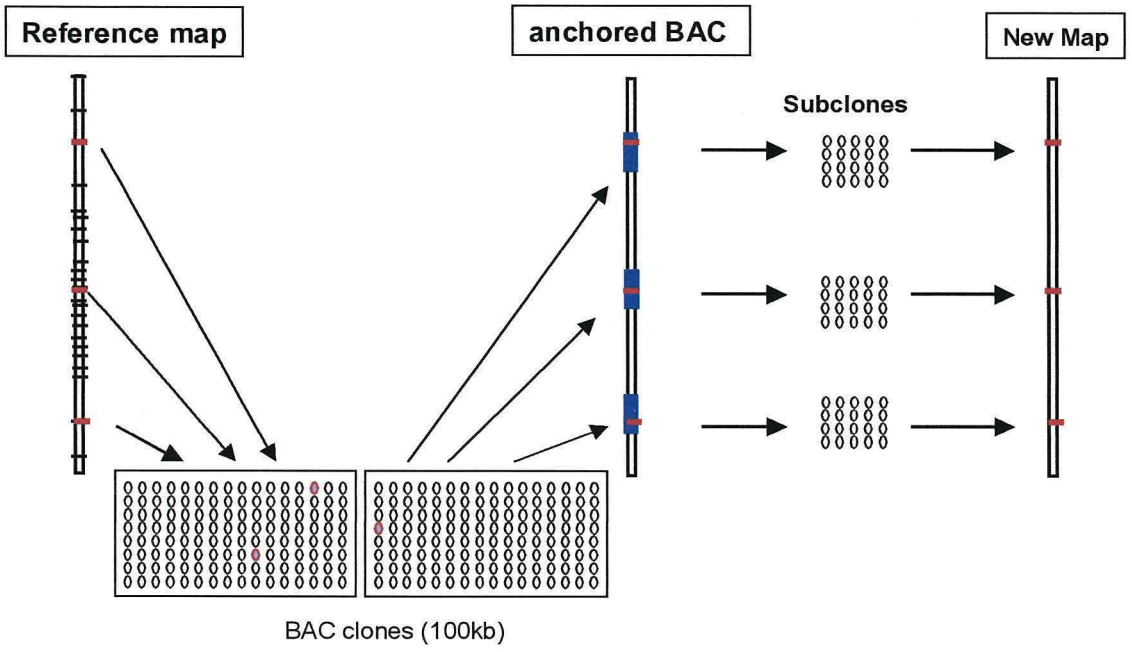


Figure 3.2: Framework maps. A set of well-spaced markers covering a genome from an existing linkage map is selected and used to screen a BAC library. A set of BAC clones covering the genome is identified. Each BAC clone is used to generate many markers (RFLP, SSR, STS) resulting in the production of a map with no clusters and no gaps.

markers had previously been mapped between these two markers in an existing map from a wide cross (Humphry, unpublished). Three of these markers were each detected by a low-copy-number probe. One of the three probes, VrCS65, detected a polymorphic marker between the PM parents and the marker was mapped in the population (Humphry *et al.* 2002). The other two probes, VrCS264 and XvrCS278, failed to detect polymorphism between the PM parents. These two monomorphic markers were used to test the feasibility of generating polymorphic markers for the targeted region in this study.

3.2 Materials and Methods

3.2.1 Plant Materials and RFLP Filters

Two mungbean genotypes, Berken and ATF-3640, and 147 recombinant inbred lines (RIL) derived from the two genotypes were used in this study. These were the parents and RIL population used for mapping genes conditioning PM resistance (Humphry, unpublished). Nylon membranes, which carry the parental DNA digested with five enzymes (*Dra*I, *Eco*RI, *Eco*RV, *Hind*III, *Hae*III) as well as the population digested with one of these enzymes were utilised in this study. These filters were prepared during construction of a mungbean linkage map in the previous study by Humphry *et al.* (2002).

3.2.2 Selection of RFLP Probes for Testing the Approach of Developing Polymorphic Markers for a Gap

Two RFLP probes, VrCS264 and VrCS278, were selected for this study. Neither of these two probes could detect polymorphism between the PM parents with the use of as many as 10 restriction enzymes (Humphry, unpublished). However, their locations in an earlier map between markers

XVrCS73 and *XVrCS296* on linkage group K (Humphry *et al.* 2002) suggested that they could provide additional markers for the chromosomal region where a major locus conditioning PM resistance was located (Figure 3.3) Inserts from these RFLP clones were used as probes to screen a BAC library (see below). These RFLP probes were prepared as described in section 2.2.12.

3.2.3 Selection of RFLP Probes Covering the Mungbean Genome

To develop a set of BAC clones covering the mungbean genome, 79 RFLP clones from three sources were used in this study (Table 3.1). Fifty-four of these were *PstI* clones from a library constructed by Humphry *et al.* (2002) and named with the prefix VrCS. Twenty-one clones form part of a skeleton map of lablab (Konduri *et al.* 2000), prefixed by LpCS, where 'Lp' represents the species name *Lablab purpureus.*

The remaining four clones were either from mungbean, named with the prefix mc and mgM, or from cowpea (*V. unguiculata* (L.) Walpers), prefixed by cgO. These mungbean and cowpea clones were a kind donation from Dr. N.D. Young, University of Minnesota, St. Paul, Minnesota, USA. All these markers were selected from the linkage map constructed by Humphry *et al.* (2002).

Two factors were considered when selecting these markers. These were the genome coverage and the copy number of a particular clone. If more than one marker was available for a given region, the one with single-copy number was chosen. This is because the BAC clones that they detected, and thus markers derived from them, would be more likely derived from, or map to, the target region.

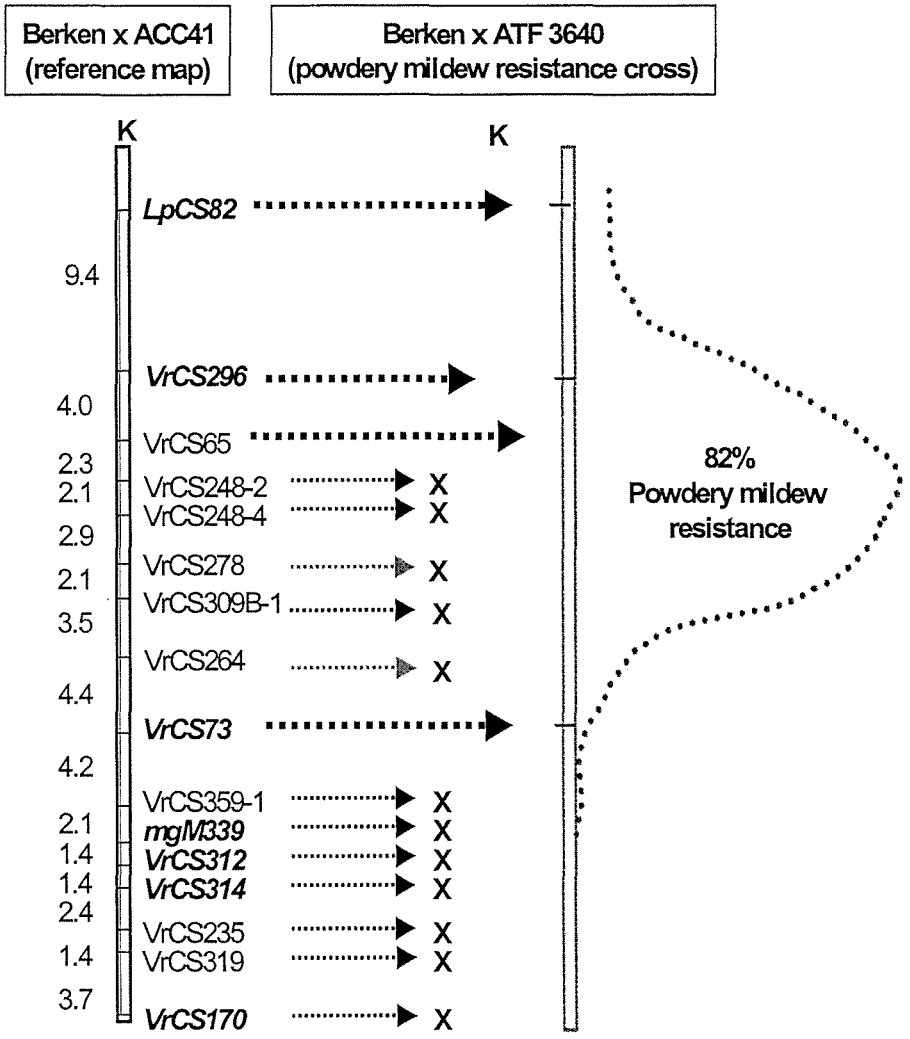


Figure 3.3: A major locus conditioning PM resistance. Markers were transferred from the reference map that was constructed using a wide cross (ACC41). Only a small proportion of the markers were mapped in the PM population (Berken x ATF-3640) and the powdery resistance gene is located in this region. VrCS264 and VrCS278 were selected from the reference map, and were used to screen the BAC library.

Table 3.1: RFLP probes covering the mungbean genome. Markers that are single copy in mungbean linkage map are indicated with **Bold** and *Italic* letters

Marker	LG	Marker	LG	Marker	LG	Marker	LG
VrCS13	D	*VrCS193*	D	*VrCS343*	B	*LpCS203*	J
VrCS17	I	*VrCS198*	E	*VrCS350*	J	*LpCS205*	F
VrCS19	H	VrCS209	B	VrCS352	M	*LpCS246*	A
VrCS23	I	*VrCS225*	B	VrCS356	A	*LpCS265*	E
VrCS49	A	*VrCS233*	B	VrCS365	L	*LpCS270*	H
VrCS53	J	VrCS246	B	*VrCS367*	A	*LpCS283*	J
VrCS65	K	*VrCS247*	D	*VrCS369*	B	*LpCS284*	J
VrCS66	E	VrCS264	K	*VrCS370*	D	*LpCS287*	F
VrCS67	A	*VrCS266*	B	VrCS371	G	*LpCS314*	I
VrCS73	K	VrCS278	K	VrCS374	F	*LpCS316*	F
VrCS84	J	VrCS282	A	*VrCS375*	L	*LpCS332*	B
VrCS85	F	*VrCS296*	K	*VrCS376*	E	*LpCS337*	A
VrCS116	H	*VrCS297*	D	*VrCS381*	G	*LpCS351*	C
VrCS122	M	VrCS309	G	*VrCS383*	D	*LpCS362*	G
VrCS126	A	*VrCS312*	K	*LpCS9*	L	LpCS405	L
VrCS150	C	*VrCS315*	D	*LpCS35*	A	*cgO026*	C
VrCS155	I	*VrCS321*	C	*LpCS54*	G	*mc004*	G
VrCS161	I	VrCS323	L	*LpCS82*	K	*mgM078*	C
VrCS170	K	*VrCS327*	G	LpCS185	L	*mgM213*	I
VrCS176	G	*VrCS342*	I	*LpCS198*	I		

All of the VrCS probes and some of LpCS probes were prepared as described in section 2.2.12. The additional steps of PCR amplification of the inserts with M13 primers, followed by phenol:chloroform:isoamylalcohol (25:24:1) and chloroform purification and then precipitation of DNA, were performed on the remaining LpCS probes and the six clones donated by Dr. N.D. Young. The PCR reaction mix was prepared in a total volume of 50 µl, consisting of 20-30 ng of plasmid, 1 x buffer, 0.125 µM dNTPs, 4 µM $MgCl_2$, 6.4 pM of M13 forward and reverse primers and 1 unit of Taq polymerase.

3.2.4 Isolation of Anchored BAC Clones to the Mungbean Linkage Map

Colony filters of 18,816 BAC clones prepared in section 2.2.11 were utilized in this study. The RFLP probes listed in Table 3.1 were used individually to screen the mungbean BAC library. Methods for probe labelling and hybridisation were described in section 2.2.13.

3.2.5 Development of Polymorphic Markers from Isolated BAC Clones

Plasmid DNA from positive BAC clones was purified and the insert sizes were determined using the methods described in section 2.2.10. Purified BAC inserts were individually digested with *Sau*3AI and purified by phenol:chloroform:isoamylalcohol (25:24:1) and chloroform extractions. The digested DNA was ligated into *Bam*HI-restricted pBluescript II SK+ vector. Fifty subclones, with insert sizes ranging from 350 bp to 900 bp, from each of the selected BAC clones were dot-blotted onto Hybond N+ nylon membranes. The subclones were probed with *Sau*3AI -restricted total genomic DNA from the mungbean genotype Berken. Those clones giving strong signals were believed to contain highly-repeated mungbean genomic sequences and were thus discarded. Inserts from the remaining putative low-copy genomic clones for each of the isolated BAC clones were

PCR-amplified using M13/M17 primers as described in section 3.2.3. Amplified low-copy subclones from each of these BACs were used as RFLP probes. They were screened individually against the PM parents with the use of five restriction enzymes (*Dra*I, *Eco*RI, *Eco*RV, *Hae*III and *Hind*III) until at least one polymorphic RFLP probe for each of the BAC clones was identified. The polymorphic probe was then used to screen the RIL population (147 RI lines). Linkage analysis was carried out using Map Manager QTX as described by Humphry *et al.* (2002).

3.3 Results

3.3.1 Developing Polymorphic Markers for Targeted Region by Exploiting BAC Clones

Two markers, VrCS264 and VrCS278, were used for this work. Screening the BAC library with VrCS264 identified a single positive BAC clone, VrBAC22N4, with an estimated insert size of 105 kbp. Thirteen of these subclones were tested as RFLP probes against the PM parents digested with restriction enzymes *Dra*I, *Eco*RI, *Eco*RV, *Hae*III and *Hind*III. Two of them, designated VrBAC22N4-A and VrBAC22N4–B, detected polymorphism in *Dra*I restriction digested and *Hae*III restriction digested PM parents, respectively (Figure 3.4). As expected, both of these probes mapped between VrCS65 and VrCS73 (Figure 3.5) [VrBAC224N-1 and -2 are called -A and -B in Figure 3.4]

Screening the BAC library with VrCS278 identified four positive clones (VrBAC20L17, VrBAC27B5, VrBAC37K3 and VrBAC47B4). Twelve of the subclones derived from VrBAC27B5 were tested against the PM parents and two of them detected polymorphism. However, none of the two markers were mapped to linkage group K. One of them mapped in linkage group B and the other in linkage group E (not shown). Similarly, testing

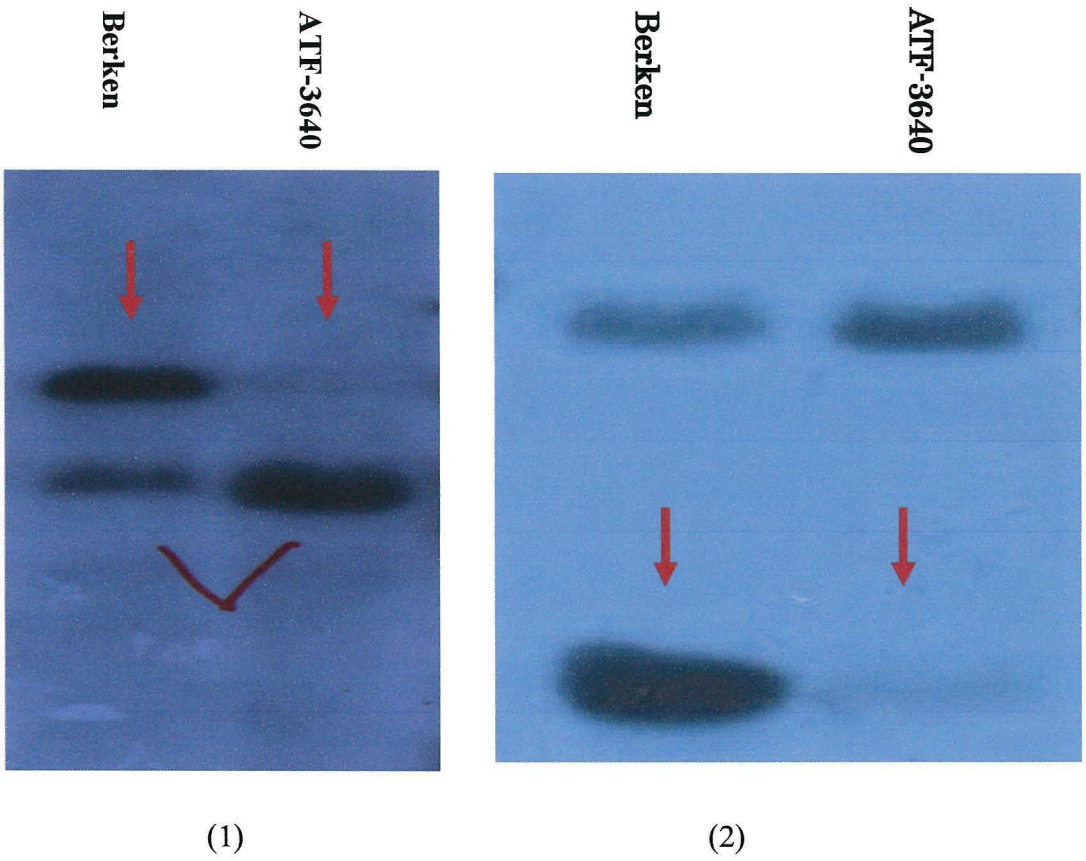


Figure 3.4: Polymorphisms in PM parents. (1) VrBAC22N4-A identified polymorphism in *Dra*I digested PM parents and (2) VrBAC22N4-B identified polymorphism in *Hae*III digested PM parents (Berken and ATF-3640)

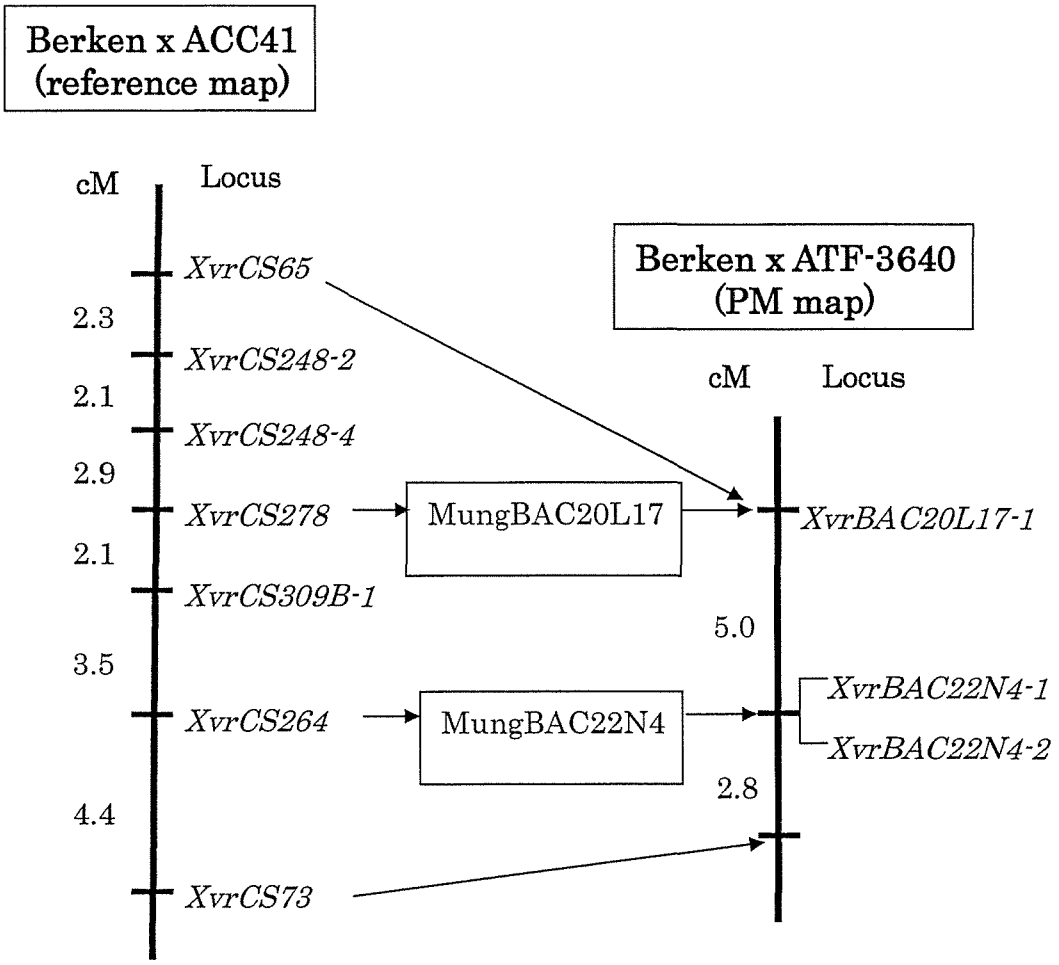


Figure 3.5: Comparative linkage maps of a section of linkage group K (where a major locus conditioning resistance to powdery mildew was located) derived from two populations, Berken x ACC41 and Berken x ATF-3640. Marker names are shown to the right and map distances in centimorgans (cM) to the left of the linkage groups that were represented by the vertical lines.

ten of the subclones derived from VrBAC37K3 identified one polymorphic marker that mapped to group J (not shown). These results suggested that these two BAC clones were not derived from the targeted genome region. Nineteen of the subclones (Figure 3.6) from VrBAC20L17 were tested against the PM parents and one of them, VrBAC20L17-1, detected polymorphism (not shown). Screening the PM population mapped the marker to the same position as XvrCS65 (Figure 3.5). Subclones from the fourth positive BAC VrBAC47B4 were not tested for polymorphic markers after the successful mapping of VrBAC20L17-1 to the expected genome region.

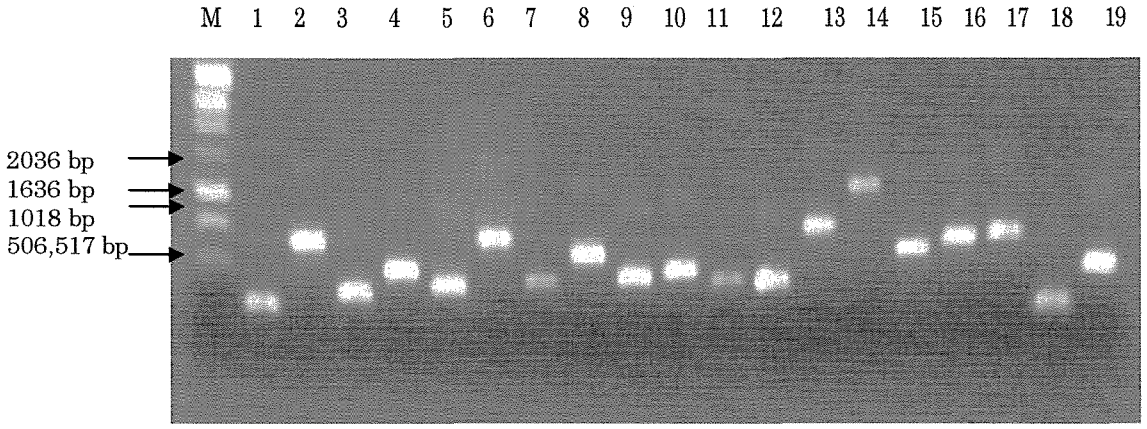


Figure 3.6: PCR amplification of VrBAC20L17 subclones. Lanes 1-19. PCR fragments amplified from representative VrBAC20L17 subclones using M13/M17 primers followed by phenol/chloroform purification. Run on 1.5% agarose gel with 1 kbp ladder (M).

3.3.2 Isolation of a Set of BAC Clones Covering the Mungbean Genome

To generate a set of BAC clones covering the mungbean genome, 79 RFLP markers were selected. These RFLP markers were used individually to screen the mungbean BAC libraries. They detected between zero to eighteen positive BAC clones, with an average of 2.5 BAC clones detected by one RFLP marker (excluding markers not used to screen the library) (Table 3.2). These markers provide good genome coverage based on their distribution on the most current linkage map (Figure 3.7).

Table 3.2: A set of anchoring BAC clones to the genetic linkage map in mungbean. Markers that are single copy in mungbean are indicated with **Bold** and *Italic* letters. Linkage group (LG) shows the position of markers in the mungbean genetic linkage map (Humphry *et al.* 2002). ' N/S' indicates that markers did not detect any BAC clone. Markers which were not tested due to technical difficulties in preparation of probes are indicated as '-'. Insert sizes are shown only where they were available.

Marker	LG	BAC clones (insert size in kbp)
VrCS13	D	8L2, 8O14, 22C3, 32N19
VrCS17	I	2P15(50), 8G20(85), 11H22(145), 11O21(50), 12B1(50), 12B20(50), 12M6(45), 27D8(80), 36A15, 37A23(90), 40H2(145), 41B17(145), 41E20(125), 42C15, 43L18, 46G9, 47O22, 48P4
VrCS19	H	2H2(110), 24O16(60), 35E16, 40G21(155), 43I7
VrCS23	I	20B22(80), 32D9, 33H16, 38C15(85), 41F14
VrCS49	A	22G3(80), 24O16(60), 38C15(85), 40H2(145), 43L18
VrCS53	J	25O22(75), 44K23
VrCS65	K	2B13(125), 6N10(120), 7F21(110), 10D2(140), 19D5(120), 19O14(120), 27B21(100), 28I14(90), 28K17(149), 31I18, 32N8, 36B23, 42J24(75), 44B1(140), 47K19(130)
VrCS66	E	47H13, 49E5
VrCS67	A	45E14
VrCS73	K	27H21(150)
VrCS84	J	20F5(70), 29C22, 32I22, 35K4, 36M17(145), 37N24(150), 41C8(100), 41G4, 42K24
VrCS85	F	N/S
VrCS116	H	34A1
VrCS122	M	22P15(55)
VrCS126	A	39G18(125), 39N21(50)
VrCS150	C	22L9, 23C17, 40D7, 40L7, 50J7(155)
VrCS155	I	12B3, 12B20(50), 27D8(80), 36A15, 41B17(150), 41E20(125), 48P4
VrCS161	I	N/S
VrCS170	K	34D1, 38E2(85), 43N13
VrCS176	G	38H14, 38J13
VrCS193	D	N/S
VrCS198	E	26A4, 31P13, 44A6
VrCS209	B	34K13, 38B14(85), 41P23, 43F8
VrCS225	B	N/S
VrCS233	B	N/S

VrCS246	B	23A7(100), 43H12
VrCS247	D	N/S
VrCS264	K	22N4(75)
VrCS266	B	25L21(85), 36I6(60)
VrCS278	K	20L17(135), 27B5(150), 37K3(50), 47B4(150)
VrCS282	A	8E16(115), 11M12(145), 25E10, 27E18(48), 29L18
VrCS296	K	23O5, 27L11, 41O2, 50K15
VrCS297	D	40D17(55)
VrCS309	G	N/S
VrCS312	K	43L12, 44K6
VrCS315	D	39B9(50)
VrCS321	C	N/S
VrCS323	L	41A13(150), 49I21, 49G4
VrCS327	G	24N15(100), 43G11
VrCS342	I	N/S
VrCS343	B	16O12(110), 26P13(50), 43N4
VrCS350	J	26K8(150)
VrCS352	M	37K18(150)
VrCS356	A	37N5(60), 37I13(150), 39O10(60)
VrCS365	L	-
VrCS367	A	41D2(45)
VrCS369	B	N/S
VrCS370	D	-
VrCS371	G	-
VrCS374	F	23L24 (60), 28F8(55), 40M18(150)
VrCS375	L	29N9, 33D20, 34J8, 37G14(195), 38L19(45)
VrCS376	E	N/S
VrCS381	G	N/S
VrCS383	D	45F11
LpCS9	L	N/S
LpCS35	A	N/S
LpCS54	G	N/S
LpCS82	K	23L24(60)
LpCS185	L	15M11(35), 15J17, 34N15(40), 39L13(160)
LpCS198	I	N/S
LpCS203	J	N/S
LpCS205	F	-

LpCS246	A	23A7(100), 43H12
LpCS265	E	23L24(60), 32D9
LpCS270	H	39L13(160)
LpCS283	J	N/S
LpCS284	J	N/S
LpCS287	F	-
LpCS314	I	N/S
LpCS316	F	30C3(45), 32M16(85), 37B14(145), 37G8(140), 37N4(135), 50J7(155)
LpCS332	B	-
LpCS337	A	-
LpCS351	C	15H4, 23P19, 33O7, 35G13, 40D11, 44K6
LpCS362	G	36D23, 44L2, 44N13, 44M15, 50D4, 50B14
LpCS405	L	8C23, 13D8, 13M19, 13I23, 25E18
cgO026	C	N/S
mc004	G	-
mgM078	C	N/S
MgM213	I	27L7(80), 29J3(135), 41E23(48), 50H14(150)

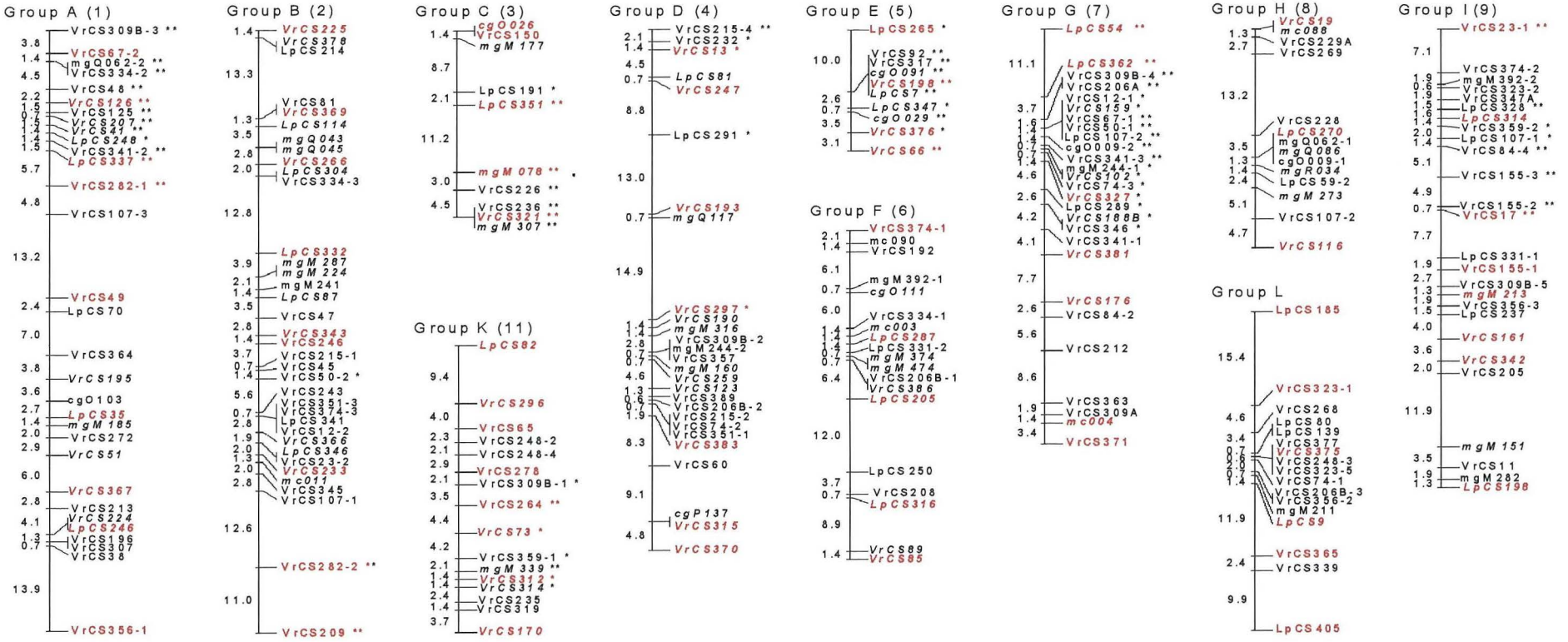


Figure 3.7: RFLP markers used to generate anchored BAC clones and their locations in a mungbean linkage map. A set of well-spaced markers (shown in red) covering a genome from an existing linkage map. The mungbean BAC library was screened against these markers to generate a set of BAC clones anchoring to the genetic linkage in mungbean. The mungbean genetic linkage map was constructed by Humphry *et al.* 2002.

3.4 Discussion

This chapter reports a new strategy that will allow efficient generation of polymorphic markers for targeted genome regions. This new strategy exploits large insert BAC clones. Due to the size differences between molecular markers (typically a few hundred base pairs) and BAC clones (100 kbp or even larger; Zhang *et al.* 1996b), a BAC clone can potentially be used to generate many molecular markers that all map to the same genome region. With multiple choices, the chance to develop a polymorphic marker from each of the identified BAC clones for any new population could be dramatically enhanced.

3.4.1 Generation of Polymorphic Markers for Specific Chromosome Region

Polymorphic markers were successfully developed for the specific chromosomal region from both of the original monomorphic markers in this study. However, compared with the use of random markers, the ratio of polymorphic probes from BAC subclones seemed to be much lower. Testing a total of 54 BAC subclones against the PM parents produced 3 markers that mapped to the target region, giving a ratio of about 5%. This figure could be as high as 10 percent (3/31) if the subclones from the false-positive BACs were not counted. Such BACs could be avoided by using single-copy-number probes to isolate BAC clones. However, this percentage is still significantly lower than the 26% obtained from the 322 RFLP markers for the same population (Humphry *et al.* 2002).

In addition to the different numbers of restriction enzymes used between the two studies, there are two other possible explanations for the lower ratio of polymorphic markers detected by the BAC subclones compared to that by random markers. Firstly, compared to other genome regions, the

target region may have an inherent low level of polymorphism. Although 26% of the 322 RFLP clones detected polymorphism between the PM parents, none of the five markers known to be located in this region (based on their map locations from an existing map (Humphry, unpublished) detected polymorphism between the PM parents. Secondly, the BAC subclones were untested random clones, but the 322 RFLP probes used were a subset of clones that had been used for generating the mungbean (Humphry, unpublished) and lablab (Konduri *et al.* 2000) linkage maps. It is possible that mapped clones, those that successfully detected polymorphism in different populations, are better able to detect polymorphism in general compared to random clones.

In contrast to the method of Michelmore *et al.* (1991) that utilises random markers, this new method exploits only markers that have the potential to be mapped in the target region. Thus it offers the potential of dramatically improving the generation of polymorphic markers for specific chromosome regions. Further, the new approach does not depend on markers flanking a target. However, it does require markers with known chromosomal locations, which may be readily available from the extensive linkage maps available for many species of agronomic importance.

3.4.2 Framework Map for New Mapping Population

The isolation of the BAC clones covering the mungbean genome is an extension of developing polymorphic markers for targeted genome regions. In contrast with using random markers that often produce linkage maps with marker clusters as well as gaps, these BAC clones offer the possibility of generating linkage maps with no marker clusters and no gaps. This is because new mapping projects can now be focused on generating one polymorphic marker for each of the genome regions by exploiting these BAC clones.

This strategy of generating polymorphic markers for targeted genome regions can readily be extended to develop a large number of highly polymorphic markers closely linked to genes of interest. Based on the location of a marker and its distance to a target, new markers on both sides of the original marker can be selected from a reference map. These markers can then be used to isolate BAC clones, and markers derived from the BAC clones at one side of the original marker would be more closely linked to the target. When combined with the method that has been successfully used to develop SSR markers from BAC clones (Cregan *et al.* 1999), the method reported here has the potential to generate closely linked and highly polymorphic markers for all those genes to which loosely linked markers have been developed, thus leading to enhanced capacity in marker-assisted breeding.

Chapter 4

Isolation of SSR and STS Markers for Targeted Genome Regions by Exploiting BAC Libraries

4.1 Introduction

To study the genetics of different characters, or to isolate markers closely linked to genes of interest, a large number of markers often need to be exploited. This is the main reason that, of the many different types of molecular markers available, RFLP and AFLP are the most widely used marker systems. During recent years, AFLP has become the dominant marker system for many mapping projects (Campbell *et al.* 2001; Potokina *et al.* 2002). As a result, markers identified so far for many genes are either RFLP or AFLP markers (Williams *et al.* 2002; Zhong and Steffenson 2002).

While RFLP and AFLP are excellent marker systems for generating linkage maps and identifying markers linked to genes of interest, they are not suitable for applications in programs of genetic studies and breeding, for several reasons. For RFLP analysis, a large quantity of DNA is essential, and radioactive labelling is often required to obtain reliable signals. For AFLP analysis, the procedure is time consuming and expensive (Smith *et al.* 2002). Further, the main advantage of AFLP is that a single analysis can generate a large number of polymorphic fragments. This makes it very different, if not impossible, to transfer mapping information from one population to another. Because of this difficulty, AFLP fragments linked to genes of interest often need to be converted to locus-specific markers for application in different programs.

A large number of genotypes are involved in a breeding program. Therefore, ideal markers for application in breeding programs should be those that are not only cheap and quick to analyse, but also locus-specific so that they can be used to follow the same genes/alleles in different genetic backgrounds. Taking these factors into consideration, PCR-based SSR (Cregan *et al.* 1999) and STS (Smith *et al.* 2002) would be two of the most suitable systems. However, sequence data are required to develop both SSR and STS markers, and, for most species, it is still cost-prohibitive to generate these markers covering a whole genome. Thus, it would be ideal if activities associated with genome mapping and marker-trait association could be carried out using AFLP, RFLP or other marker systems which allow rapid identification of large numbers of polymorphic markers, and then SSR and STS markers could be developed for targeted genomic regions. In fact, converting RFLP or AFLP markers into SSR or STS markers has been actively pursued by many researchers (Bouzidi *et al.* 2002; Meksem *et al.* 2001; Wang *et al.* 2002; Zhong and Steffenson 2002).

The routine procedure for converting a RFLP or AFLP marker into a locus-specific and PCR-based marker usually involves the cloning and sequencing of the target DNA clone (for RFLP) or fragments (for AFLP) (Bouzidi *et al.* 2002; Zhong and Steffenson 2002). This means that a single STS can be obtained from a single RFLP clone or a single AFLP fragment. Due to the limited level of polymorphism in many species, the chance for a single STS to detect a difference between a pair of genotypes can be low (Liu *et al.* 1999). Therefore, it would be ideal if several STS or SSR markers could be generated from a single marker, with the SSR and STS markers all retaining a similar linkage to the target location as the initial marker. Several STS and SSR markers for a given gene would dramatically enhance the chance of breeders or geneticists to detect polymorphism between a given pair of genotypes.

With the availability of large insert clones, it has now become feasible to isolate a large number of STS and SSRs markers from an RFLP or AFLP marker. This feasibility was demonstrated using YAC (Chen *et al.* 1995) and BAC (Cregan *et al.* 1999). This study aimed to generate several SSR and STS markers for a major gene conditioning bruchid resistance by exploiting the mungbean BAC libraries that have been constructed (see Chapter 2).

Bruchid beetles are a serious problem in many crop species. In mungbean studies, they are often referred to simply as 'seed weevils' or 'bruchids'. These pests belong to the genus *Callosobruchus* (*Coleoptera: Bruchidae*). Two of the most damaging species are *C. chinesis* and *C. maculatus*. These insects attack stored seeds and can lead to nearly complete loss (Talekar 1988). Currently, no resistant cultivar is available, and the insects are controlled by the use of insecticides (Talekar 1988). Application of insecticides to control bruchids, however, adds extra cost to growers. Further, residuals of the insecticides are also a serious concern for the mungbean industry. For these reasons, bruchid resistance has been one of the main breeding objectives in mungbean improvement. To effectively breed bruchid resistant cultivar, sources of bruchid resistance have been actively sought (Fuji *et al.* 1989), and inheritance of the resistance has been studied (Young *et al.* 1992). As part of the mungbean breeding effort at CSIRO Plant Industry, sources of bruchid resistance were identified and genetic studies have located a major gene conditioning resistance on linkage group I. Several RFLP markers closely linked to the gene have been isolated (Humphry *et al.* 2002). One of these markers, mgM213, was selected for this study because it is a single-copy clone that is easier to manipulate compared with non-single copy sequences.

4.2 Materials and Methods

4.2.1 BAC Library Screening and Subcloning of BACs

Colony filters containing 18816 mungbean BAC clones constructed in section 2.2.11 were utilised in this study. A locus conditioning bruchid resistance is located on linkage group I (Figure 4.1). A single-copy RFLP marker, mgM213, closely linked to the gene (Humphry *et al.* 2002) was used to screen the whole mungbean BAC library to identify BAC clones containing mgM213 sequences. Positive BAC clones identified were then used as a source of DNA for the construction of small-insert libraries suitable for screening SSR markers. Each BAC clone was subcloned as described in section 3.2.5 except that *Bam*HI and *Sau*3AI restriction endonucleases were used in this study.

4.2.2 Screening for the Presence of (AT)n and (ATT)n SSRs

Transformants derived from each BAC clone with each enzyme digestion were manually picked into 384-well microtiter and 96-well microtitre plates containing LB/ampicillin (0.1mg/ml). Two sets of colony filters containing these transformants were prepared as described in section 2.2.11 except ampicillin was used as the selective agent.

Colony hybridization was performed using the protocol of Cregan *et al.* (1994). Each set of filters were placed in a box containing 6 x SSC and washed for 5 minutes at room temperature followed by 2 hours pre-wash [0.05 M Tris-HCL (pH8.0), 1.0 M NaCL, 0.001 M EDTA, 0.1% SDS] at 42°C with a change of buffer each hour. The final wash (0.1 SSC, 0.5% SDS) was done for one hour at 65°C and membranes were then transferred to pre-hybridisation buffer (6 x SSPE, 5x Denhardt's III, and 1% SDS) for two hours at 37°C. Membranes were removed from the

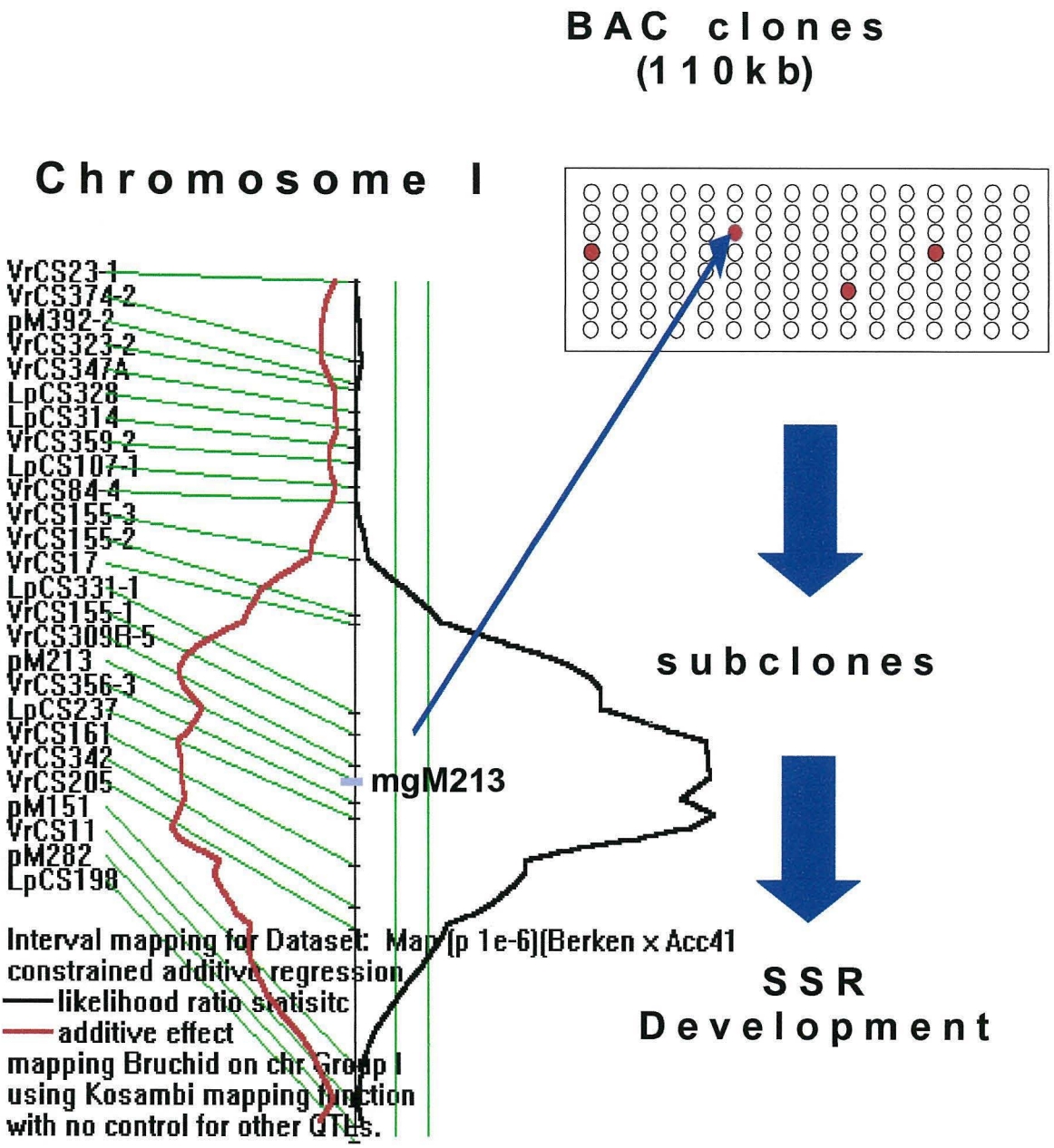


Figure 4.1: A locus conditioning bruchid resistance, genetic linkage group I (Humphry *et al.* 2002). BAC library was screened with mgM213 marker and identified BAC clones containing mgM213 sequence. Each BAC clone was subcloned for SSR development.

pre-hybridisation buffer and rinsed with hybridization buffer (6x SSPE and 1% SDS) to remove excess Denhardt's. Membranes were transferred to hybridization buffer and denatured oligonucleotide probes (Proligo, USA), $(At)_{15}CC$ or $(ATT)_{10}CCC$ labelled with α-^{32}PdATP (Perkin-Elmer, USA) were added. Hybridisation was performed at 38 °C overnight.

Following hybridization, membranes were washed in 1 x SSC with 0.1% SDS for 1 hour with a change of washing buffer each half an hour at 38 °C. Membranes were blotted dry and exposed to X-ray film for three days.

4.2.3 Sequencing Positive Subclones

Plasmids of all positive subclones identified with oligonucleotide probing were isolated by the alkaline lysis method and resuspended in 40 µl water. Plasmid DNA was treated with RNase followed by phenol:chloroform:isoamylalcohol (25:24:1) and chloroform purification. Plasmid was then precipitated and resuspended in 10 µl of water. DNA concentration was quantified on a gel. In a 0.2 ml tube, 300-500 ng plasmid DNA was transferred and water was added to make up to 6 µl plus 3.2 pM of either T3 or T7 primers and 6 µl of BigDye (Applied Biosystems). Plasmid DNA was amplified with 25 cycles under the following conditions: 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes.

When the reaction was finished, the reaction mix was transferred into a 1.5 ml tube and 7 µl of water was added to make up to 20 µl. Plasmid was precipitated with 4 µl of 3 M sodium acetate (pH5.3) and 50 µl of 100% ice-cold ethanol. The contents were briefly mixed and placed in a freezer (-20°C) for 30 minutes. Tubes were centrifuged for 30 minutes at 4°C and pellets were washed in ice-cold 70% ethanol by centrifuging for 15 minutes.

The purified inserts were sequenced using the ABI Prism dRhodamine Terminater Cycle Sequencing Kit (Perkin-Elmer, Applied Biosystems, Foster City, California). Inserts were sequenced from both ends and consensus sequences were determined by analysis with the Sequencer™ 3.1.1 (Gene Code, Ann Arbor, MI, USA) software. After consensus sequences were determined, all sequences from positive subclones were compared using the software in order to detect duplicate sequences. The sequences which contained a dinucleotide repeat of 20 or more base pairs, or a trinucleotide repeat of 21 or more base pairs, were considered to possess an SSR (Cregan *et al.* 1999)

4.2.4 Designing Primers and Testing

For each unique microsatellite that was identified, and for which sufficient reliable flanking sequence data were available, optimal oligonucleotide primers to the flanking regions were selected using the primer selection program Sequencer™ 3.1.1 (Gene Code, Ann Arbor, MI, USA). STS primers were also designed for other subclones, which were identified using either $(AT)_{15}CC$ or $(ATT)_{10}CCC$ oligonucleotide probes but which did not possess an SSR.

4.3 Results

4.3.1 Isolation of BAC Clones Containing the mgM213 Sequence

Of the several RFLP markers closely linked to a major gene conditioning bruchid resistance (Humphry, unpublished), marker mgM213 was selected for this work because it is a single copy clone. Screening the 18,816 mungbean BAC clones using the RFLP clone mgM213 identified four positive BAC clones. The names of these BAC clones and their insert sizes are listed in Table 4.1.

Table 4.1: BAC clones containing mgM213 sequences and their sizes

Name	Insert size (kbp)
VrBAC 27L7	80
VrBAC 29J13	135
VrBAC 41E23	48
VrBAC 50H14	155

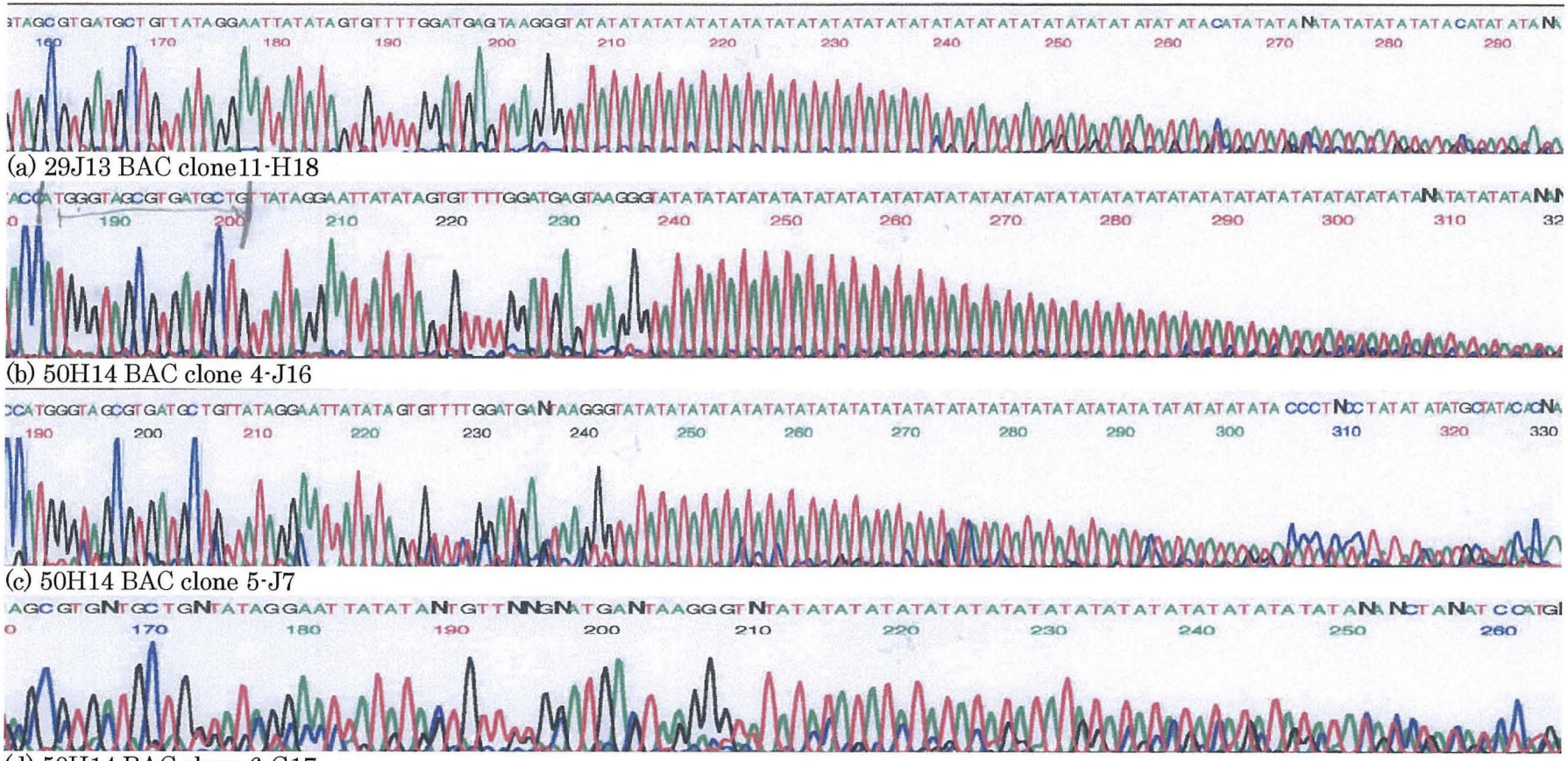
4.3.2 Isolation of BAC Subclones that Contain SSR Sequences

Two of these positive BAC clones, VrBAC41E23 and VrBAC50H14 were subcloned using *Bam*HI and *Sau*3AI restriction enzymes while VrBAC27L7 and VrBAC29J13 were subcloned using only *Sau*3AI. Subclones generated from each BAC clone and/or different restriction endonucleases were picked individually into 384-well microtitre plates and 96-well microtitre plates. The number of subclones picked for each BAC clone and for each restriction enzyme is shown in Table 4.2.

Table 4.2: Number of subclones picked to prepare colony filters for each BAC clone

	*Bam*HI	*Sau*3AI
VrBAC 27L7	n/a	1152 (3x384)
VrBAC 29J13	n/a	1152 (3x384)
VrBAC 41E23	1152 (3x384)	1248 (3x384+1x96)
VrBAC 50H14	1152 (3x384)	2400 (6x384+1x96)

Colony filters for these subclones were screened with oligonucleotide probes, $(AT)_{15}CC$ and $(ATT)_{10}CCC$, labelled using α-^{32}PdATP. Eight positive subclones were identified by $(AT)_{15}CC$ and two subclones from the $(ATT)_{10}CCC$ probe. All of these positive BAC clones were sequenced. As shown in Figure 4.2 four of eight subclones identified using the $(AT)_{15}CC$ probe contained enough (AT) repeats to warrant their SSR marker status.



(a) 29J13 BAC clone11-H18

(b) 50H14 BAC clone 4-J16

(c) 50H14 BAC clone 5-J7

(d) 50H14 BAC clone 6-G17

Figure 4.2: SSR sequences identified using $(AT)_{15}CC$ probe in four different subclones derived from two BAC clones (29J13, 135 kbp and 50H14, 155 kbp). All four SSRs were found to be consensus.

However, upon further analysis, they were found to be contigs. Simple repeats contained by other subclones were too short to be treated as SSR markers (see Discussion).

4.3.3 Designing Primers for SSR and STS Markers

Based on the sequences flanking the AT repeats, a pair of primers were designed for analysing this SSR marker. Similarly, three pairs of STS primers were designed for 3 positive subclones identified using either $(AT)_{15}CC$ or $(ATT)_{10}CCC$ oligonucleotide probes. These primer sequences are shown in Table 4.3.

4.3.4 SSR and STS Analysis

Due to time constraints, the oligonucleotide primers that had been designed were not tested on either the two cultivars, ACC41 and Berken, or their RI lines (see Discussion).

Table 4.3: SSR and STS primer sequences used to detect polymorphism between ACC41 and Berken

Marker	BAC clone (enzyme)	Primer pair	Primer sequence
SSR	50H14(Sau3AI)	MicroF (18-mer)	5'-ATGGGTAGCGTGATGCTG-3'
		MicroR (19-mer)	5'-TGTCAAAATGTGGTTGGCG-3'
STS	47E23(BamHI)	sts1F (22-mer)	5'-GATTTGAGTTTCTATGCCCTCC-3'
		sts1R (19-mer)	5'-CTCTTCCCTGATTCCTCGC-3'
	50H14(BamHI)	sts2F (24-mer)	5'-AACTGCCAGAAACTACCAGGTGTC-3'
		sts2R (22-mer)	5'-AGAGGAGGGGAAGAGAATCACG-3'
	50H14(Sau3AI)	sts3F (19-mer)	5'-CCCAACCTTATTACAAGCC-3'
		sts3R (19-mer)	5'-TCAACCTTCATCAACACCC-3'

4.4 Discussion

Developing markers for a given character is invariably carried out with a limited number of genotypes and progenies. This is because hundreds or even thousands of markers are often needed to identify a marker closely linked to a gene of interest, and it can be very costly and time consuming to go through the large numbers of markers. On the contrary, only one or a few markers are needed to follow a particular gene but hundreds or even thousands of genotypes are routinely required to be screened in a breeding program. Clearly, markers suitable for mapping projects are not necessarily suitable for breeding programs. Ideal markers for genome mapping are those that can generate many polymorphic markers from a single analysis. AFLP fits into this category and that is why it has become so popular (Katengam *et al.* 2002; Miftahudin *et al.* 2002). Ideal markers for a breeding program should be those that are locus-specific and can be analysed using a simple procedure. This is the reason why SSR (Cregan *et al.* 1999) and STS (Smith *et al.* 2002) are useful for this purpose.

Markers closely linked to genes conditioning many characters have been identified. However, application of these markers in breeding programs is still limited due to the reason that many of the markers are not suitable for this application. Thus, converting RFLP and AFLP markers into PCR-based markers that can be effectively used in breeding programs is now becoming an international focus. Naturally, SSR (Cregan *et al.* 1999) and STS (Smith *et al.* 2002) have become the systems of choice for breeding programs due to their locus specific nature and their simple analysis by standard PCR reactions.

4.4.1 BAC Libraries as Intermediaries for Developing PCR-Based and Locus-Specific Markers

There are two popular approaches for converting markers into PCR-based and locus-specific markers. The first one is by sequencing the target (clones for RFLP and fragments for AFLP) and then designing primers based on the sequence data (Smith *et al.* 2002). This approach normally produces a single STS marker from each initial clone or fragment.

The second approach is via large insert clones, such as YAC or BAC. This approach exploits the size difference between a marker (usually a few hundred base pairs) and a large insert clone (usually 100,000 base pairs or larger). Thus a single large insert clone can be converted into many markers. Further, due to its large size, a large insert clone can not only be used to generate STS, but also other marker types such as SSRs (Chen *et al.* 1995, Cregan *et al.* 1994), which has been proved to be the most polymorphic marker system (Cregan *et al.* 1999). For these reasons, the BAC clones that had been constructed for developing STS and SSR markers were chosen to develop markers for the locus conditioning bruchid resistance.

4.4.2 Presence of (AT)n and (ATT)n SSRs in Mungbean

SSRs or microsatellites are 2-5 nucleotide core units that are tandemly repeated in the genome (Litt and Luty 1989). Of the many different core units, (AT)n and (ATT)n have been found to be most abundant in legumes (Chen *et al.* 1995; Cregan *et al.* 1999). Thus these two oligonucleotides were selected for developing SSR markers in mungbean in this work.

Screening all the subclones with $(AT)_{15}CC$ oligonucleotides revealed that possibly eight of them contained this repeat sequence. However, when the

eight subclones were sequenced only four of them were proven to contain the repeat sequence. Further analysis discovered that these four sequences were identical (Figure 4.2).

$(ATT)_{10}CCC$ generated only two positive subclones and they both produced very strong hybridization signals. However, sequencing of these positive subclones failed to find trinucleotide ATT repeats of more than seven, the minimum length widely used to define a microsatellite or SSR (Beckman and Weber 1992, Wang *et al.* 1994).

4.4.3 Frequency of SSR Sequences in Mungbean

The largest BAC clone (50H14) used for this work is 155 kbp (Table 4.1). Surveying results indicated that on average one SSR can be found in every 21.2 kbp legume DNA (Wang *et al.* 1994). Thus, the single SSR identified from the four BAC clones is well under the expected number of SSRs that could potentially be found. This, however, does not seem to be the norm in mungbean. For example, four (AT)n SSRs and three (ATT)n SSRs have been isolated from two BAC clones for another locus conditioning powdery mildew resistance (Humphry, unpublished), using a similar approach.

Of the many different marker systems available, SSR or microsatellite possesses several features which make it the marker of choice for breeding programs. Firstly, SSRs are the most polymorphic marker systems available and have the ability to detect multiple alleles for a given locus (Cregan *et al.* 1999). This feature is very important because lack of polymorphism is a common problem in marker application; SSRs are a co-dominant marker system; SSR sequences are widespread in eukaryotic genomes (Wang *et al.* 1994); and finally, the procedure and techniques for SSR analysis are simple (Cregan *et al.* 1999).

4.4.4 PCR Amplification Analysis

In this study, one SSR and three STS markers for a major gene conditioning bruchid resistance in mungbean were successfully identified by exploiting mungbean BAC clones. These user-friendly markers would have the potential to enhance the progress of incorporating bruchid resistance into commercial mungbean cultivars in breeding programs. They will also facilitate efforts of fine mapping of the resistance gene, which will be important for further characterization and manipulation of the gene for use, not only in mungbean, but also in other related grain crop species.

Due to the time constraints of this project, the primer pairs designed in section 4.3.3 were not able to be tested to detect polymorphism in the Berken x ACC41 mapping population. Analysis of the primer sequences suggest that an amplification using the primers designed would yield PCR products of 300 to 350 bp. Future testing of primers, given a project which allowed such work to be done, would prove their reliability.

Chapter 5

General Discussion

Genome research has become a key international activity. This activity has not only provided deep insights into genome structure (see Gale and Devos 1998b) and evolution (Tanksley *et al.* 1988), but it has also offered exciting new opportunities for developing superior new plant and animal genotypes. To this end, genome research has been focused on marker-aided breeding and map-based gene cloning.

Detailed molecular marker maps have been developed for most plant and animal species. However, current technologies are slow and expensive to generate linkage maps. It often takes a small team many months to produce a linkage map, and most maps produced to date contain gaps (Hayden *et al.* 2001). These gaps limit the power of these linkage maps in locating genes/QTLs because they are powerless to detect those genes/QTLs located in those gaps.

Similarly, the current technologies are also slow to generate markers that are useful in a breeding program. It often takes years to find one or two markers linked to a targeted gene/QTL. These markers have only limited usefulness in breeding programs. This is because the ability for a marker to detect differences between any given pair of genotypes is usually low, but hundreds or even thousands genotypes are used in a breeding program. Thus, to be able to reveal differences between most of the genotypes, many markers are needed for a given gene/QTL. This is, unfortunately, not the case for most of the target genes/QTLs. The results presented in the preceding chapters demonstrated that some of the major limitations in marker projects could be dramatically alleviated by exploiting BAC technologies.

5.1 Efficient Generation of Polymorphic Markers for Gaps in Linkage Maps

Gaps are common in linkage maps constructed to date, and they limit the power of maps in locating genes conditioning a trait and in isolating markers closely linked to genes of interest. The only effective method of filling gaps is the BSA analysis. This method requires flanking markers for a target so the approach cannot be used for those gaps located at ends of linkage groups (Campbell *et al.* 2001). Further, BSA utilizes random markers. Thus it can be very time-consuming to get polymorphic markers for some gaps.

The results from this study demonstrated that it is feasible to isolate polymorphic markers for targeted gaps in linkage maps by exploiting BAC clones. This new strategy has two major advantages compared to BSA. Firstly, it does not require flanking markers and therefore it can be used for gaps in any sections of a linkage map. Secondly, it exploits only markers that have a good chance to fall into the gaps, thus it could dramatically improve the efficiency of gap filling, which in turn offers great potential to dramatically enhance the efficiency of locating genes and isolating markers closely linked to genes of interest.

5.2 Development of Locus Specific SSR and STS Markers for Targeted Genome Regions by Exploiting BAC Technology

In addition to the considerations of speed and cost, the procedures of marker analysis will have to be easy and safe to be able to find practical application in breeding programs. This is because large numbers of genotypes are required in a breeding program. For the same reason, a single marker would not be adequate to follow a specific gene/QTL because it can only detect difference between a small proportion of the genotypes

needed for a breeding program. Thus, RFLP or AFLP markers that linked to genes/QTLs of interest would need to be converted to PCR-based markers. Ideally several PCR-based markers could be developed from a single marker linked to the gene. Using a RFLP marker linked to a major gene conditioning bruchid resistance, it was demonstrated in this study that it is possible to generate a large number of PCR based markers from a single RFLP marker if BAC clones are available. In addition, the SSR and STS markers generated from this study can be very useful in incorporating the bruchid resistance into the current breeding programs of mungbean and related species.

5.3 Efficient Generation of Framework Maps by Exploiting BAC Technologies

A large number of markers are often required for a mapping project. This is one of the reasons why it can be very time consuming to construct a linkage map. However, it is important to understand that the number of markers is not important in a mapping program. Rather, the distribution and locations of markers are more important. For example, a marker every 20 cM would be adequate in mapping QTLs in a population consisting of a few hundred individuals (Tanksley 1993). What this means is that mapping does not have to be so time-consuming if only markers that are located in selected regions of a genome are used.

To develop a capacity for effectively generating framework maps, a set of BAC clones covering the mungbean genome was isolated (see Chapter 3). As demonstrated in Chapter 3, it is not difficult to generate polymorphic markers when BAC clones are available. Thus, further mapping projects in mungbean can be carried out by developing a polymorphic marker from each of the anchored BAC clones. This approach will make it possible to produce framework maps with good genome coverage and few gaps – thus

allowing more efficient genome mapping and marker identification.

5.4 General Conclusion

Though numerous molecular marker systems are available, constructing a linkage map with good genome coverage is still difficult. The difficulty is reflected by gaps that can be common in linkage maps. These gaps limit the effectiveness of a whole genome scan to identify genes/genomic regions underlying a trait. A method based on comparing two pooled DNA samples has been widely used to isolate polymorphic markers for gaps. In this study, an alternative method that allows more efficient isolation of polymorphic markers for gaps is reported. This alternative method is based on a combined exploitation of molecular markers from reference maps and BAC libraries that are available for many species. DNA markers from reference maps can be used to isolate BAC clones that originate from specific genomic regions. Due to its large insert, a single BAC clone can potentially be converted into a large number of molecular markers that would all map to the same genome region. With such a large number of markers available, the chance of developing a polymorphic marker from each BAC clone for any new population would be dramatically enhanced. The feasibility of this new strategy was demonstrated in mungbean (*Vigna radiata*) using a chromosome region harbouring a major locus conditioning powdery mildew (PM) resistance as the target. This approach to isolating polymorphic markers for gaps in linkage maps can be readily extended to generate DNA markers more closely linked to genes of interest.

Of the many marker types available to date, locus-specific SSR or microsatellites are known to be the most informative marker system since the system can be easily assayed by standard PCR technology, and is highly transferable between populations. A mungbean BAC library

constructed in this study was also utilized to develop SSR markers closely linked to a major locus conditioning resistance of bruchid, one of the most serious problems of mungbean and other *Vigna* species. The SSR markers will not only facilitate MAS but also the detailed mapping of the locus.

The large inserts BAC clones anchored to the genetic linkage map in mungbean developed in this study not only provide an immediate substrate for further analyses, but they also present a resource for construction of a high-resolution map in the region of interest.

References

Abbo S., Miller T.E. and King L.P. (1993) Primer-induced *in situ* hybridisation to plant chromosomes. *Genome* 36 (5): 815-817.

Albertsen H.M., Abderrahim H., Cann H.M., Dausset J., Paslier D.J. and Cohen D. (1990) Construction and characterisation of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4256-4260.

Aledo R., Raz R., Monfort A., Vincient C.M., Puigdomenech P. and Martinez-Izquierdo J.A. (1995) Chromosome localisation and characterisation of a family of long interspersed repetitive DNA elements from the genus *Zea*. *Theoretical and Applied Genetics* 90: 1094-1100.

Anderson C. (1993) Genome shortcut leads to problems. *Science* 259: 1684-1687.

Ardiel G.S., Grewal T.S., Deberdt P., Rossnagel B.G. and Scoles G.J. (2002) Inheritance of resistance to covered smut in barley and development of a tightly linked SCAR marker. *Theoretical and Applied Genetics* 104: 457-464.

Arondel V., Lemieux B., Hwang I., Gibson S., Goodman H.M. and Somerville C.R. (1992) Map-based cloning of gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. *Science* 258: 1353-1355.

Arumuganathan K. and Earle E.D. (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9: 208-218.

Asnaghi C., Paulet F., Kaye G., Grivet L., Deu M., Glaszmann J.C. and D'Hont A. (2000) Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics* 101: 962-969.

Barua U.M., Chalmers K.J., Hackett C.A., Thomas W.T.B., Powell W. and Waugh R. (1993) Identification of RAPD markers linked to a *Rhynchosporium secalis* resistant locus in barley using near-isogenic lines and bulked segregation analysis. *Heredity* 71: 177-184.

Barzen E., Mechelke W., Ritter E., Schulte-Kappert E. and Salamini F. (1995) An extended map of the sugar beet genome, containing RFLP and RAPD loci. *Theoretical and Applied Genetics* 90: 189-193.

Baudracco-Arnas S. and Pitrat M. (1996) A genetic map of melon (*Cucumis melo* L.) with RFLP, RAPD, isozyme, disease resistance and morphological markers. *Theoretical and Applied Genetics* 93: 57-64.

Beavis W.D. and Grant D. (1991) A linkage map based on information from four F2 population of maize (*Zea mays* L.). *Theoretical and Applied Genetics* 82: 636-644.

Beckman J.S. and Weber J.L. (1992) Survey of human and rat microsatellites. *Genomics* 12: 627-631.

Bent A.F., Kunkel B.N., Dahlbeck D., Brown K.L., Schmidt R., Giraudat J., Leung J. and Staskawicz B.J. (1994) *RPS2* of *Arabidopsis thaliana*, a leucine rich repeat class of plant disease resistance gene. *Science* 265: 1856-1860.

Bonierbale M.W., Plaisted R.L. and Tanksley S.D. (1988) RFLP maps based on common sets of clones reveal modes of chromosome evolution in potato and tomato. *Genetics* 120: 1095-1203.

Botstein D., White R.L., Skolnik M. and Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314-331.

Bouzidi M.F., Dadaoui S., Cambon F., Vear F., Tourvielle De Labrouhe D., Nicolas P. and Mouzeyar S. (2002) Molecular analysis of a major locus for resistance to downy mildew in sunflower with specific PCR-based markers. *Theoretical and Applied Genetics* 104: 592-600.

Buerstmayr H., Lemmens M., Hartl L., Doldi L., Steiner B., Stierschneider M. and Ruckenbauer P. (2002) Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). *Theoretical and Applied Genetics* 104: 84-91.

Burke D.T., Carle G.F. and Olson M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806-811.

Burr B., Burr F.A., Thompson K.H., Albertson M.C. and Stuber C.W. (1988) Gene mapping with recombinant inbreds in maize. *Genetics* 118: 519-526.

Caetano-Anolles G., Bassam B.J. and Gresshoff P.M. (1991) DNA amplification of fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology* 9: 553-556.

Campbell A.W., Daggard G., Bekes F., Pedler A., Sutherland M.W. and Appels R. (2001) Targeting AFLP-DNA markers to specific traits and chromosome regions. *Australian Journal of Agricultural Research* 52: 1153-1160.

Chalmers K.J., Campbell A.W., Kretschmer J., Karakousis A., Henschke P.H., Pierens S., Harker N., Pallotta M., Cornish G.B., Shariflou M.R., Rampling L.R., MacLauchlan A., Daggard G., Sharp P.J., Holton T.A., Sutherland M.W., Apples R. and Langridge P. (2001) Construction of three linkage maps in bread wheat, *Triticum aestivum. Australian Journal of Agricultural Research* 52: 1089-1119.

Chang C., Bowman J.L., DeJohn A.W., Lander E.S. and Meyerowitz E.M. (1988) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America* 85: 6856-6860.

Chao S., Sharp P.J., Worland A.J., Warham E.J., Koebner R.M.D. and Gale M.D. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theoretical and Applied Genetics* 78: 495-504.

Chen H., Pulido J.C. and Duyk G.M. (1995) MATS: a rapid and efficient method for the development of microsatellite markers from YACs. *Genomics* 25: 1-8.

Cheung W.Y., Friesen L., Rakow G.F.W., Seguin-Swartz G. and Landry B.S. (1997) A RFLP-based linkage map of mustard [*B. juncea* (L.) Czern and Coss]. *Theoretical and Applied Genetics* 94: 841-851.

Choi S., Creelman R., Mullet J.E. and Wing R.A. (1995) Construction and characterisation of a bacterial artificial chromosome library of *Arabidopsis thaliana. Plant Molecular Biology Reporter* 13: 124-128.

Choi S. and Wing R.A. (2000) The construction of bacterial artificial chromosome (BAC) libraries.
http://www.genome.clemson.edu/ protocols2new.html

Chumakov I.M., Rigault P., Gall I.L., Bellanné-Chantelot C., Billault A., Guillou S., Soularue P., Guasconi G., Poullier E., Gras I., Belova M., Sambucy J.L., Susini L., Gervy G., Gilbert F., Beaufils S., Bui H., Massart C., De Tand M.F., Dukasz F., Lecoulant S., Ougen P., Perrot V., Saumier M., Soravito C., Bahouayila R., Cohen-Akenine A., Barillot E., Bertrand S., Codani J.J., Caterina D., Georges I., Lacroix B., Lucotte G., Sahbatou M., Schmid C., Sangouard M., Tubacher E., Dib C., Fauré S., Fizames C., Gyapay G., Millasseau P., Nguyen S., Muselet D., Vignal A., Morissette J., Menninger J., Lieman J., Desai T., Banks A., Bray-Ward P., Ward D., Hudson T., Gerety S., Toote S., Stein L., Page D.C., Lander E.S., Weissenbach J., Paslier D.L. and Cohen D. (1995) A YAC contig map of the human genome. *Nature* 377: 175-298.

Chyi Y.S., Hoenecke M.E. and Sernyk J.L. (1992) A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. Campestris). *Genome* 35: 746-757.

Civardi L., Xia Y.J.U., Edwarts K.J., Schnable P.S. and Nikolau B.J. (1994) The relationship between genetic and physical distances in the cloned a al-sh2 interval of the *Zea mays* L. genome. *Proceedings of the National Academy of Sciences of the United States of America* 91: 8268-8272.

Clarke L. and Carbon J. (1976) A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9: 91-100.

Collins J. and Hohn B. (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proceedings of the National Academy of Sciences of the United States of America* 75: 4242-4246.

Cordeiro G.M., Casu R., McIntyre C.L., Manners J.M. and Henry R.J. (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Science* 160: 1115-1123.

Cregan P.B., Bhagwat A.A., Akkaya M.S. and Rongwen J. (1994) Microsatellite Fingerprinting and Mapping of Soybean. *Methods in Molecular and Cellular Biology* 5: 49-61.

Cregan P.B., Mudge J., Fickus E.W., Marek L.F., Danesh D., Denny R., Shoemaker R.C., Matthews B.F., Jarvik T. and Young N.D. (1999) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theoretical and Applied Genetics* 98: 919-928.

Danesh D., Penuela S., Mudge J., Denny R.L., Norstrom H., Martinez J.P. and Young N.D. (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. *Theoretical and Applied Genetics* 96: 196-202.

Eagles H.A., Bariana H.S., Ogbonnaya F.C., Rebetzke G.J., Hollamby G.J., Henry R.J., Henschke P.H. and Carter M. (2001) Implementation of markers in Australian wheat breeding. *Australian Journal of Agricultural Research* 52: 1349-1356.

Ecker J.R. (1990) PFGE YAC analysis of the *Arabidopsis* genome. *Methods* 1: 186-194.

Edwards K.J., Thompson H., Edwards D., Saizieu A.D., Sparks C., Thompson J.A., Greenland A.J., Eyers M. and Schuch W. (1992) Construction and characterisation of a yeast artificial chromosome library containing three haploid maize genome equivalents. *Plant Molecular Biology* 19: 299-308.

Edwards M.D., Stuber C.W. and Wendel J.F. (1987) Molecular facilitated investigation of quantitative trait loci in maize. I Numbers, genomic distribution and types of gene action. *Genetics* 116: 113-125.

Foisset N., Delourme R., Barret P., Hubert N., Landry B.S. and Renard M. (1996) Molecular mapping analysis in *B. napus* using isozyme, RAPD and RFLP markers on a double haploid progeny. *Theoretical and Applied Genetics* 93: 1017-1025.

Fregene M., Angel F., Gomez R., Rodriguez F., Chavarriaga P., Roca W., Tohme J. and Bonierbale M. (1997) A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics* 95: 431-441.

Frijters A.C.J., Zhang Z., van Damm M., Wang G.L., Ronald P.C. and Michelmore R.W. (1997) Construction of a BAC library containing large *Eco*RI and *Hind*III genomic fragments of lettuce. *Theoretical and Applied Genetics* 94: 390-399.

Fuji K., Ishimoto M. and Kitamura K. (1989) Patterns of resistance to bean weevils (Bruchidae) in *Vigna radiata* – sublobata complex inform the breeding of new resistant varieties. *Applied Entomology and Zoology* 24: 126-132.

Gale M.D. and Devos K.M. (1998a) Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* 95: 1071-1974.

Gale M.D. and Devos K.M. (1998b) Plant comparative genetics after 10 years. *Science* 282: 656-658.

Gale M.D. and Witcombe J.R. (1992) *DNA markers and marker-mediated applications in plant breeding, with particular reference to pearl millet breeding.* In: Biotechnology and crop improvement in Asia (ed. JP Moss) pp 323-332 ICRISAT, Patancheru, A.P, India.

Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschabel U., Kaufmann H., Thampson R.D., Bonierbale M.W., Ganal M.W., Tanksley S.D. and Salamini F. (1991) RFLP maps of potato and their alignment with the homologous tomato genome. *Theoretical and Applied Genetics* 83: 6419-6423.

Gill K.S., Lubbers E.L., Gill B.S., Raupp W.J. and Cox T.S. (1991) A genetic linkage map of *T. tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). *Genome* 34: 362-374.

Gilpin B.J., McCallum J.A., Frew T.J. and Timmerman-Vaughan G.M. (1997) A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). *Theoretical and Applied Genetics* 95: 1289-1299.

Goodman M.M. and Stuber C.W. (1983) *Maize* In: Isozymes in plant genetics and breeding, part B (eds. Tanksley SD and Orton T) pp 1-33 Elsevier Scientific Publishing Co., Amsterdam.

Grandillo S. and Tanksley S.D. (1996) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium. Theoretical and Applied Genetics* 92: 957-965.

Grill E. and Somerville C. (1991) Construction and characterisation of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Molecular and General Genetics* 226: 484-490.

Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166: 557-580.

Hart G.E. (1983) *Hexaploid wheat (Triticum aestivum L. Em Thell)* In: Isozymes in plant genetics and breeding, part B (eds Tanksley SD and Orton T) pp35-56 Elsevier Scientific Publishing Co., Amsterdam.

Harushima Y., Yano M., Shomura A., Satao M., Shimano T., Kuboki Y., Yamamoto T., Lin S.Y., Antonio B.A., Parco A., Kajiya H., Huang N., Yamamoto K., Nagamura Y., Kurata N., Khush G. and Sasaki T. (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148 (1): 479-494.

Hayden M.J., Khatkar S. and Sharp P.J. (2001) Targetting microsatellites (SSRs) in genetic linkage maps of bread wheat. *Australian Journal of Agricultural Research* 52: 1143-1152.

Hohn B. and Collings J. (1988) Ten years of cosmids. *Trends in Biotechnology.* 6: 293-298.

http://www.genome.clemson.edu

http://www.tamu.edu

Humphry M.E., Konduri V., Lambrides C.J., Magner T., McIntyre C.L. and Liu C.J. (2002) Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of synteny between the two genomes. *Theoretical and Applied Genetics* 105: 160-166.

Iannou P.A., Amemiya C.T., Granes J., Kroisel P.M., Shizuya H., Chen C., Batzer M.A. and de Jong P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics* 6: 84-89.

Jacobs H.J., Lindpainter K., Lincoln S.E., Kusumi K., Bunker R.K., Mao Y.P., Dzau Ganten D. and Lander E.S. (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypersensitive rat. *Cell* 67: 213-224.

Jellen E.N., Gill B.C. and Cox T.S. (1994) Genomic *in situ* hybridisation differentiates between A/D-genome and C-genome chromatin and detects intergenomic translocations in polyploid oat species (genus *Avena*). *Genome* 37: 613-618.

Jones N., Ougham H. and Thomas H. (1997) Markers and mapping: we are all geneticists now. *New Phytology* 137: 165-177.

Kanazawa H., Noumi T. and Futai M. (1986) Analysis of *Escherichia coli* of the H+-transporting ATPase: Determination of altered site of the structural genes. *Methods in Enzymology* 126: 595-603.

Katengam S., Crane J.M. and Knapp S.J. (2002) The development of a genetic map for meadowfoam comprised of amplified fragment length polymorphism. *Theoretical and Applied Genetics* 104: 92-96

Kearsey M.J. (1997) *Genetic resources and plant breeding (identification, mapping and manipulation of simple and complex traits)* In: Biotechnology and Plant genetic resources, Conservation and Use (eds Callow JA, Ford-Lloyd BV and Newbury HJ) pp175-202 CAB International, UK.

Keim P., Schupp J.M., Travis S.E., Clayton K., Zhu T., Shi L., Ferreira A. and Webb D.M. (1997) A high density soybean genetic map based on AFLP markers. *Crop Science* 37: 537-543.

Kesseli R.V., Paran I. and Michelmore R.W. (1994) Analysis of detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics* 136: 1435-1446.

Kiehne K and Neale D (1998) DNA pooling strategy for saturation mapping in outbred crosses. *Molecular Breeding* 4: 179-185

King I.P., Purdie K.A., Orford S.E., Reader A.M. and Miller T.E. (1993) Detection of homologous chiasma formation in *Triticum durum* x *Thinopyrum bessarabicum* hybrids using genomic *in situ* hybridisation. *Heredity* 71: 369-372.

Koch J., Kolvraa S., Petersen K.B., Gregersen N. and Bolund L. (1989) Oligonucleotide-priming methods for the chromosome specific labelling of alpha satellite DNA *in situ*. *Chromosoma* 98: 259-265.

Konduri V., Godwin I.D. and Liu C.J. (2000) Genetic mapping of the *Lablab purpureus* genome suggests the presence of 'cuckoo' gene(s) in this species. *Theoretical and Applied Genetics* 100: 866-871.

Kota R., Wolf M., Michalek W. and Graner A. (2001) Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.). *Genome* 44: 523-528

Kumer L.S. (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143-182.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E. and Newburg L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.

Landry B.S., Hubert N., Crete R., Chiang M.S., Lincoln S.E. and Etoh T. (1992) A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). *Genome* 35: 409-420.

Laptian N.V.L., Sears R.G., Rayburn A.L. and Gill B.S. (1986) Wheat-rye translocations. *Journal of Heredity* 77: 415-419.

Laurie D.A. (1988) Annual Report Institute of Plant Science Research/John Innes Institute pp3-4.

Laurie D.A., Bryan G.J. and Snape J.W. (1997) *Genomic Relationships, Conserved Synteny and Wide-hybrids.* In: Biotechnology and Plant genetic resources, Conservation and Use (eds Callow JA, Ford-Llyd BV and Newbury HJ) pp77-101 CAB International, UK.

Le H.T., Armstrong K. and Miki B. (1989) Detection of rye DNA in wheat-rye hybrids and wheat translocation stocks using total genomic DNA as probe. *Plant Molecular Biology Reporter* 7: 150-158.

Lefebvre V. and Chèvre A.M. (1995) Tools for marking plant disease and pest resistance genes: a review. *Agronomies* 15: 3-19.

Libert F., Lefort A., Okimito R., Womack J. and Georgies M. (1993) Construction of a bovin genomic library of large yeast artificial chromosome clones. *Genomics* 18: 270-276.

Lippman Z. and Tanksley S.D. (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. asculentum* var. Giant Heirloom. *Genetics* 158:413-422.

Litt M. and Luty J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44: 397-401.

Liu C.J. (1996) Genetic diversity and relationships among *Lablab purpureus* genotypes evaluated using random amplified polymorphic DNAs as markers. *Euphytica* 90: 115-119.

Liu C.J. (1997) *Rapid biochemical and molecular techniques in genetic purity tests.* In: Asian seed '97, technical report 2 (20).

Liu C.J. and Musial J.M. (1997) *Stylosanthes* sp. aff. *Scabra*: a putative diploid progenitor of *Stylosanthes scabra*. *Plant Systematics and Evolution* 208: 99-105.

Liu C.J. and Musial J.M. (2001) The application of chloroplast DNA clones in identifying maternal donors for polyploid species of Stylosanthes. *Theoretical and Applied Genetics* 102: 73-77.

Liu C.J., Musial J.M. and Smith F.W. (1996) Evidence for a low level of genomic specificity of sequence-tagged-sites in Stylosanthes. *Theoretical and Applied Genetics* 88: 864-868.

Liu C.J., Musial J.M. and Thomas B.D. (1999) Genetic relationships among Stylosanthes species based on RFLP and STS analyses. *Theoretical and Applied Genetics* 99: 1179-1186.

Liu C.J., Witcombe J.R., Pittaway T.S., Nash M., Hash C.T., Busso C.S. and Gale M.D. (1994) An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). *Theoretical and Applied Genetics* 89: 481-487.

Liu W.Z., Biyashev R.M. and Saghai-Maroof M.A. (1996) Development of SSR DNA markers and their integration into a barley linkage map. *Theoretical and Applied Genetics* 93: 869-876.

Loarce Y., Hueros G. and Ferrer E. (1996) A molecular linkage map of rye. *Theoretical and Applied Genetics* 93: 1112-1118.

Luo M., Wang Y-H., Frisch D., Joobeur T., Wing R.A. and Dean R.A. (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*). *Genome* 44: 154-162.

Marek L.F. and Shoemaker R.C. (1997) BAC contig development by fingerprint analysis in soybean. *Genome* 40: 420-427.

Markert C.L. and Moller F. (1959) Multiple forms of enzymes: tissues, oncogenic, and species specific patterns. *Proceedings of the National Academy of Sciences of the United States of America* 45: 753-763.

Martin G.B., Brommonschenkel S.H., Chunwongse J., Frary A., Ganal M.W., Spivey R., Wu T., Earle E.D. and Tanksley S.D. (1993) Map-based cloning of protein Kinase gene conferring disease resistance in tomato. *Science* 262: 1432-1436.

Martin G.B., Ganal M.W. and Tanksley S.D. (1992) Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Molecular and General Genetics* 233: 25-32.

McClintock B. (1929) Chromosome morphology of *Zea may. Science* 69: 629-630.

McGrath J.M. and Quiros C.F. (1991) Inheritance of isozyme and RFLP markers in *Brassica campestris* and comparison with *B. oleracea. Theoretical and Applied Genetics* 82: 668-673.

Meksem K., Ruben E., Hyten D., Triwitayakorn K. and Lightfoot D.A. (2001) Conversion of AFLP bands into high-throughput DNA markers. *Molecular Genetic and Genomics:* 265: 207-214.

Menancio-Hautea D., Fatokun C.A., Kumar L., Danesh D. and Young N.D. (1993a) Comparative genome analysis of mungbean (*V. radiata* L. Wilczek) and cowpea (*V. unguiculata* L. Walpers) using RFLP mapping data. *Theoretical and Applied Genetics* 86: 797-810.

Menancio-Hautea D., Kumar L., Danesh D. and Young N.D. (1993b) *A genome, map for mungbean [Vigna radiata (L.) Wilczek] based on DNA genetic markers (2n=2x=22).* pp6259-6261. In: O'Brien, J.S. editor. Genetic Maps: A Compilation of Linkage and Restriction Maps of Genetically Studied Organisms. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Menedez C.M., Hall A.E. and Gepts P. (1997) A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. *Theoretical and Applied Genetics* 95: 1210-1217.

Messmer M.M., Keller M., Zanetti S. and Keller B. (1999) Genetic linkage map of a wheat x spelt cross. *Theoretical and Applied Genetics* 98:1163-1170.

Michelmore R.W., Paran I. and Kesseli R.V. (1991) Identification of markers linked to disease resistance genes by bulked-segregation analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America* 88: 9828-9832.

Miftahudin, Scoles G.J., and Gustafson J.P. (2002) AFLP markers tightly linked to the aluminium-tolerance gene *Alt3* in rye (*Secale cereale* L.). *Theoretical and Applied Genetics* 104: 626-631.

Monaco A.P. and Larin Z. (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends in Biotechnology* 12: 280-286.

Money T.A., Liu C.J. and Gale M.D. (1994) *Conversion of RFLP markers for downy mildew resistance in pearl millet to sequence-tagged-sites.* pp 65-68. In: Witcombe, R.J. and Duncan R.R. (eds). Use of molecular

markers in sorghum and pearl millet breeding for developing countries. Proc ODA Plant Sci Res Conf, Oversead Development Administration, London, U.K.

Moore G., Foote T., Helentjaris T., Devos K., Kurata N. and Gale M. (1995) Was there a single ancestral cereal chromosome? *Trends in Genetics* 11: 81-82.

Mozo T., Fischer S., Shizuya H. and Altmann T. (1998) Construction and characterisation of the IGF *Arabidopsis* BAC library. *Molecular and General Genetics* 258: 562-570.

Mudge J., Andersen W.R., Kehrer R.L. and Fairbanks D.J. (1996) A RAPD genetic map of *Saccharum officinarum*. *Crop Science* 36: 1362-1366.

Mukai Y., Endo T.R. and Gill B.S. (1991) Physical mapping of the 18S.26S rRNA multigene family in common wheat: identification of a new locus. *Chromosoma* 100: 71-78.

Mukai Y. and Gill B.S. (1991) Detection of barley chromatin added to wheat by genomic *in situ* hybridisation. *Genome* 34: 448-452.

Mukai Y., Nakahara Y. and Yamamoto M. (1993) Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence *in situ* hybridisation using total genomic and highly repeated DNA probes. *Genome* 36: 489-494.

Mullis K.B. and Faloona F.A. (1987) Specific synthesis of DNA *in vitro* via polymerase-catalysed chain reaction. *Methods in Enzymology* 155: 335-350.

Nakamura S., Asakawa S., Ohmido N., Fukui K., Shimizu N. and Kawasaki S. (1997) Construction of an 800 kb contig in the near-centromeric region of the rice blast resistance gene *Pi-ta* using a highly representative rice BAC library. *Molecular and General Genetics* 254: 611-620.

Nelson J.C., Sorrells M.E., Van Deynze A.E., Lu Y.H., Atkinson M., Bernard M., Leroy P., Faris J. and Anderson J. (1995) Molecular mapping of wheat. Major genes and rearrangements in homologous group 4, 5 and 7. *Genetics* 141: 721-731.

Nowatny P., Kwon J.M. and Goate A. (2001) SNP analysis to dissect human traits. *Current Opinion in Neurobiology* 11: 637-641.

O'Donoughue L.S., Kianian S.F., Rayapati R.J., Penner G.A., Sorrells M.E., Tanksley S.D., Philips R.L., Rines R.L., Lee M., Fedak G., Moinar S.J., Hoffman D., Salas C.A., Wu B., Autrique E. and Van Deynze A. (1995) A molecular linkage map of cultivated oat. *Genome* 38: 368-380.

Orita M., Iwahana H., Kanazawa H., Hayashi K. and Sekiya T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* 86: 2766-2770.

Oslon M., Hood L., Cantor C. and Doststein D. (1989). A common language for physical mapping of the human genome. *Science* 254: 1434-1435.

Osoegawa K., Woon P.Y., Zhao B., Frengen E., Tateno M., Ctanese J.J. and de Jong P. J. (1998) An Improvement Approach for Construction of Bacterial Artificial Chromosome Libraries. *Genomics* 52: 1-8.

Paglia G.P., Olivieria M. and Morgante M. (1998) Towards second-generation STS (sequence-tagged sites) linkage maps in conifers: a genetic map of Norway spruce (*Picea abies* K). *Molecular and General Genetics* 258: 466-478.

Paillard M., Lashermes P. and Petiard V. (1996) Construction of a molecular linkage map in coffee. *Theoretical and Applied Genetics* 93: 41-47.

Paran I. and Michelmore R.W. (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* 85: 985-993.

Paterson A.H., Lin Y.R., Li Z., Schertz K.F., Doebley J.F., Pinson S.R.M., Liu S.C., Stansel J.W. and Irvine J.E. (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714-1718.

Paterson A.H., Tanksley S.D. and Sorrells M.E. (1991) DNA markers in plant improvement. *Advance in Agronomy* 46: 40-90.

Penner G.A., Bush A., Wise R., Kim W., Domier L., Kasha K., Laroche A., Scoles G., Molnar S.J. and Fedek G. (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Application* 2: 341-345.

Potokina E., Blattner F.R., Alexandrova T. and Bachmann K. (2002) AFLP diversity in the common vetch (*Vicia sativa L.*) on the world scale. *Theoretical and Applied Genetics* 105: 58-67.

Qualset C.O., Schaller C.W. and Williams J.C. (1965) Performance of isogenic line of barley as influenced by awn length, linkage blocks and environment. *Crop Science* 5: 489-494.

Rafalski J.A. (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science* 162 :329-333.

Ratnaparkhe M.B., Santra D.K., Tullu A. and Muehlbauer F.J. (1998) Inheritance of inter-simple sequence repeat polymorphisms and linkage with a fusarium wilt resistance gene in chickpea. *Theoretical and Applied Genetics* 96: 348-353.

Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M-H., Leroy P. and Ganal M.W. (1998) A microsatellite map of wheat. *Genetics* 149: 2007-2023.

Roy J.K., Prasad M., Varshney R.K., Balyan H.S., Blake T.K., Dhaliwal H.S., Singh H., Edwards K.J. and Gupta P.K. (1999) Identification of a microsatellite on chromosomes 6B and a STS on 7D of bread wheat showing an association with preharvest sprouting tolerance. *Theoretical and Applied Genetics* 99: 336-340.

Russel R.J. (1990) Genetics (second edition) pp.913. Scott, Foresman and Company: Glenview, Illinois USA 99: 419-424.

Saiki R.K., Gelfand D.H., Stoffel S., Schars S.J., Higuchi R., Horn G.T., Mullis K.B. and Erlich H.A. (1988) Primer-detected enzymatic amplification of DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 8: 1477-1481.

Salimath S.S., de Oliveira A.C., Godwin I.D. and Bennetzer J.L. (1995) Assessment of genomic orihins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome* 38: 757-763.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

San Miguel P., Tikhonov A., Jun Y-K., Motochoulskaia N., Akharov D., Melake-Berhan A., Springer P.S., Edward K.J., Avramova Z. and Bennetzen J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.

Schoen D.J. and Brown A.H.D. (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the National Academy of Sciences of the United States of America* 90: 10623-10627.

Schulz-Schaeffer J. (1980) Cytogenetics. Plants. Animals and humans. pp446 Springer-Verlag: New York.

Sessions S.K. (1990) *Chromosome: molecular cytogenetics.* In: Hillis, D.M. and Moritz, C (eds) Molecular systematics. Sinauer Associates, Sunderland, MA, pp156-203.

Shappley Z.W., Jenkins J.N., Watson C.E. Jr., Kahler A.H. and Meredith W.R. (1996) Establishment of molecular markers and linkage groups in two F2 populations of upland cotton. *Theoretical and Applied Genetics* 92: 915-919.

Sharp P.J., Kreis M., Shewry P.R. and Gale M.D. (1988) Location of beta-amylase sequences in wheat and its relatives. *Theoretical and Applied Genetics* 75: 286-290.

Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y. and Simon M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-base vector. *Proceedings of the National Academy of Sciences of the United States of America* 89: 8794-8797.

Smith P.H., Koebner R.M.D. and Boyd L.A. (2002) The development of a STS marker linked to a yellow rust resistance derived from the wheat cultivar Moro. *Theoretical and Applied Genetics* 104: 1278-1282.

Soller M. and Beckmann J.S. (1990) Marker-based mapping of quantitative trait loci using replicated progenies. *Theoretical and Applied Genetics* 80: 205-208.

Sondur S.N., Manshardt R.M. and Stiles J.I. (1996) A genetic linkage map of papaya based on RAPD markers. *Theoretical and Applied Genetics* 93: 547-553.

Stam P. (1993) Construction of integrated genetic linkage maps by means of a computer package Joinmap. *The Plant Journal* 5: 739-744.

Stuber C.W. (1992) Biochemical and molecular markers in plant breeding. *Plant Breeding Review* 9: 37-61.

Stuber C.W., Edwards M.D. and Wendel J.F. (1987) Molecular marker facilitated investigation of quantitative trait loci in maize. II Factors influencing yield and its component trait. *Crop Science* 27: 639-648.

Summers R.W., Koebner R.M.D., Hollins T.W., Forster J. and Macartney D.P. (1988) *The use of isozyme marker in breeding wheat (Triticum aestivum) resistance to the eyespot pathogen (Pseudocercosporella herpotrichoides).* Proceedings of the 7th International Wheat Genetics Symposium pp 1195-1197.

Talekar N.S. (1988) Biology, damage and control of bruchid pests of mungbean. In: Mungbean : Proceedings of the Second International Symposium, Asian Vegetable Research and Development Center, Shanhua, Taiwan, pp329-342.

Tanksley S.D. (1983) Molecular markers in plant breeding. *Plant Molecular Biology Reporter* 1 (1): 3-8.

Tanksley S.D. (1993) Mapping polygenes. *Annual Review in Genetics* 27: 205-233.

Tanksley S.D., Bernatzky R., Lapitan N.L. and Prince J.P. (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Sciences of the United States of America* 85: 6419-6423.

Tanksley S.D., Ganal M.W. and Martin G.B. (1995) Chromosome landing: a paragram for map-based gene cloning in plants with large genomes. *Trends in Genetics* 2:63-68.

Tanksley S.D., Ganal M.W., Prince J.P., de Vincente M.C., Bonierbale M.W., Broun P., Fulton T.M., Giovannoini J.J., Grandillo S., Martin G.B., Messeguer R., Miller J.C., Miller L., Paterson A.H., Pineda O., Rodes M.S., Wing R.A., Wu W. and Young N.D. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genome* 132: 1141-1160.

Tanksley S.D. and Rick C.M. (1980) Isozymic gene linkage map of the tomato: Applications in genetics and breeding. *Theoretical and Applied Genetics* 57: 161-170.

Tanksley S.D., Young N.D., Paterson A.H. and Bonierbale M.W. (1989) RFLP mapping in plant breeding; New tools for an old Science. *Biotechnology* 7: 257-264.

Tao Q. and Zhang H-B. (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. *Nucleic Acids Research* 26 (21): 4901-4909.

Tao Y.Z., Jordan D.R., Henzell R.G. and McIntyre C.L. (1998) Construction of genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps. *Australian Journal of Agricultural Research* 49: 726-736.

Thumma B.R., Naidu B.P., Chandra A., Cameron D.F., Bahnisch L.M. and Liu C.J. (2001) Identification of causal relationships among traits related to drought resistance in *Stylosanthes scabra* using QTL analysis. *Journal of Experimental Botany* 52: 203-214.

Tomkins J.P., Mahalingham R., Miller-Smith H., Goicoechea J.L., Knapp H.T. and Wing R.A. (1999a) A soybean bacterial artificial chromosome library for PI 437654 and the identification of clones associated with cyst nematode resistance. *Plant Molecular Biology.* 41: 25-32.

Tomkins J.P., Yu Y., Miller-Smith H., Frisch D.A., Woo S. and Wing R.A. (1999b) A bacterial artificial chromosome library for sugar cane. *Theoretical and Applied Genetics* 99: 419-424.

Trask B.J. (1991) Fluorescence *in situ* hybridisation: application in cytogenetics and gene mapping. *Trends in Genetics* 7: 149-154.

Ulloa M., Meredith W.R. Jr., Shappley Z.W. and Kahler A.L. (2002) RFLP genetic linkage maps from four F2.3 populations and a joinmap of *Gossypium hirsutum* L. *Theoretical and Applied Genetics* 104: 200-208.

Umehara Y., Inagaki A., Tanoue H., Hayashiguchi Y., Nagamura Y., Saji S., Otsuki Y., Fujimura T., Kurata N. and Minobe Y. (1995) Construction and characterization of a rice YAC library for physical mapping. *Molecular Breeding* 1: 79-89.

Vallejos C.E., Sakiyama N.S. and Chase C.D. (1992) A molecular marker-based linkage map of *Phaseolus vulgaris. Genetics* 131: 733-740.

van Eck H.J., van der Voort J.R., Draaistra J., van Zandvoort P., van Enckevort E., Segers B., Peleman J., Jacobsen E., Helder J. and Bakker J. (1995) The inheritance and chromosomal localisation of AFLP markers in a non-inbred potato offspring. *Molecular Breeding* 1: 397-410.

Vega J.M., Abbo S., Feldman M. and Levy A.A. (1994) Chromosome painting in plants: *in situ* hybridisation with a DNA probe from a specific microdissected chromosome arm of common wheat. *Proceedings of the National Academy of Sciences of the United States of America* 91: 12041-12045.

Vinatzer B.A., Zhang H-B. and Sansavini S. (1998) Construction and characterisation of a bacterial artificial chromosome library of apple. *Theoretical and Applied Genetics* 97: 1183-1190.

Vodenicharova M. (1989) Use of proteins as molecular-genetic markers in plant. *Genetic Selection* 22: 269-277.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Freitjers A., Pot J., Peleman J., Kuiper M. and Zebeau M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 409-411.

Wang G.L., Holsten T.E., Song W.Y., Wang H.P. and Ronald P.C. (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. *Plant Journal* 7: 525-533.

Wang Y., Georgi L.L., Zhebentyayeva T.N., Reighard G.L., Scorza R. and Abbott A.G. (2002) High-throughput targeted SSR marker development in peach (*Prunus persica). Genome* 45: 319-328.

Wang Z., Weber J.L., Zhang G. and Tanksley S.D. (1994) Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics* 88: 1-6.

Ward E.R. and Jen G.C. (1990) Isolation of single-copy-sequence clones from yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. *Plant Molecular Biology* 14: 561-568.

Weeden N.R., Muehlbauer F.J. and Ladizinsky G. (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of Heredity* 83: 123-129.

Welsh J. and McClelland M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18 (22): 7213-7218.

Westman A.L. and Kresovich S. (1997) *Use of Molecular Marker Techniques for Detection of Plant Genetic Variation.* In: Biotechnology and Plant genetic resources, Conservation and Use (eds Callow JA, Ford-Lloyd BV and Newbury HJ) pp9-48 CAB International, UK.

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A. and Tingey S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18 (22): 6351-6535.

Williams K.J., Taylor S.P., Bogacki P., Pallotta M., Bariana H.S. and Wallwork H. (2002) Mapping of the root lesion nematode (*Pratylenchus neglectus*) resistance gene *Rlnn1* in wheat. *Theoretical and Applied Genetics* 104: 874-879.

Woo S.S., Jiang J., Gill B.S., Paterson A.H. and Wing R.A. (1994) Construction and characterisation of a bacterial artificial chromosome library of *Sorghum bicolour*. *Nucleic Acids Research* 22: 4299-4931.

Woo S.S., Rastogi V.K., Zhang H.B., Paterson A.H., Schertz K.F. and Wing R.A. (1995) Isolation of Megabase-Size DNA from Sorghum and Application for physical Mapping and Bacterial and Yeast Artificial Chromosome Library Construction. *Plant Molecular Biology Reporter* 13 (1): 82-94.

Xu G.W., Magill C.W., Schertz K.F. and Hart G.E. (1994) A RFLP linkage map of *Sorghum bicolor* (L.) Moench. *Theoretical and Applied Genetics* 89: 139-145.

Yoshimura S., Umehara Y., Kurata N., Nagamura Y., Sasaki T., Minobe Y. and Iwata N. (1996) Identification of a YAC clone carrying the Xa-1 allele, a bacterial blight resistance gene in rice. *Theoretical and Applied Genetics* 93:117-122.

Young N.D., Kumar L., Menancio-Hautea D., Danesh D., Talekar N.S., Shanmugasundarum S. and Kim D.H. (1992) RFLP mapping of a major bruchid resistance gene in mungbean (*Vigna radiata*, L. Wilczek). *Theoretical and Applied Genetics* 84: 839-844.

Zabeau M. and Vos P. (1994) Selective restriction fragment amplification: a general method for DNA fingerprinting. *European Patent Application* 92402629.7 (publication number:0534858 A1).

Zhang H.B., Choi S., Woo S.S., Li Z. and Wing R.A. (1996a) Construction and characterisation of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. *Molecular Breeding* 2: 11-24.

Zhang H.B., Martin G.B., Tanksley S.D. and Wing R.A. (1994) Map-based cloning in crop plants. Tomato as a model system: II. Isolation and characterisation of a set of overlapping yeast artificial chromosomes encompassing the jointless locus. *Molecular and General Genetics* 244: 613-621.

Zhang H.B. and Tao Q. (1998) Cloning and stable maintenance of DNA fragments over 300kb in *Escherichia coli* with conventional plasmid-based vectors. *Nucleic Acids Research* 26: 4091-4909.

Zhang H.B. and Wing R.A. (1997) Physical mapping of the rice genome with BACs. *Plant Molecular Biology* 35: 115-127.

Zhang H.B., Woo S.S. and Wing R.A. (1996b) *BAC, YAC and Cosmid library construction.* In: Plant Gene Isolation: Principles and Practice. Foster, G.D. and Twell, D. (Eds) John Wiley & Sons Ltd. England pp 75-99.

Zhang H.B. and Wu C. (2001) BAC as tools for genome sequencing. *Plant Physiology and Biochemistry* 39: 195-209.

Zhang H-B., Zhao X., Ding X., Paterson A.H. and Wing R.A. (1995) Preparation of megabase-size DNA from plant nuclei. *The Plant Journal* 7(1): 175-184.

Zhang Q., Shen B.Z., Dai X.K., Mei M.H., Saghai-Maroof M.A. and Li Z.B. (1994) Using bulked extremes and recessive class to map genes for photo-period-sensitive genic male sterility in rice. *Proceedings of the National Academy of Sciences of the United States of America* 94: 8675-8679.

Zhong S. and Steffenson B.J. (2002) Identification and characterization of DNA markers associated with a locus conferring virulence on barley in the plant pathogenic fungus *Cochliobolus sativus. Theoretical and Applied Genetics* 104:1049-1054.

Zietkiewcz E., Rafalski A. and Labuda D. (1994) Genome fingerprinting by simple sequence repeats (SSR) anchored PCR amplification. *Genomics* 20: 176-183.